DOCUMENT RESUME

ED 169 498 CS 004 762

AUTHOR Bhola, H. S., Ed.
TITLE Evaluating Functional Literacy.
INSTITUTION International Inst. for Adult Literacy Methods, Teheran (Iran).
PUB DATE 79
NOTE 177p.; Part of the "Literacy in Development: A Series of Training Monographs"

EDRS PRICE MF01/PC08 Plus Postage.
DESCRIPTORS *Adult Basic Education; Evaluation Criteria; *Evaluation Methods; Foreign Countries; *Functional Illiteracy; *Literacy Education; Models; *Program Evaluation

ABSTRACT

Written primarily for use in training, this monograph is addressed to the middle-level evaluation worker in literacy projects and programs. The 12 chapters deal with the following topics: evaluation as a link in the "ideology to technology" chain, evaluation models and approaches, the situation-specific strategy model of evaluation, making evaluation operational in functional literacy programs, the classical protocols in the measurement of change, some rediscovered protocols in the measurement of change, implementation data for management and evaluation, special evaluation tasks for functional literacy workers, inspecting and analyzing data, evaluation design, and interpreting evaluation results for decision making. A selected annotated bibliography and a glossary of terms are appended. (PL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED169498

Literacy in development: a series of training monographs
Series editor, H. S. Bhola, Indiana University

Evaluating functional literacy

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

H. S. Bhola

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
IIALM

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Hulton Educational Publications Ltd.,
in co-operation with the
International Institute for Adult Literacy Methods
Tehran 1979

CS004762

© IIALM 1979
ISBN 0 7175 0815 3

Published in 1979 by Hulton Educational Publications Ltd., Raans Road, Amersham, Bucks., in co-operation with the International Institute for Adult Literacy Methods, P.O. Box 1555, Tehran, Iran.

Text set in 12/13 pt VIP Times, printed by photolithography, and bound in Great Britain at The Pitman Press, Bath.

Invitation to the reader

Reading a book can be like conversing with a knowledgeable friend. But it is a one-way conversation. The author speaks and his readers listen. Frequently, the author anticipates his reader's thoughts and answers his questions. Otherwise, these questions remain unanswered.

In order to assist the readers of this monograph, the International Institute for Adult Literacy Methods (IIALM) seeks to engage him in a dialogue. Today, we live in an age in which communications are rapid and not expensive. Therefore, we suggest that the reader send his questions to the IIALM, as well as any problems which might arise in the use of this monograph. If the Institute itself cannot help, it will endeavour to put the reader in touch with someone who can.

We want to know if you, the reader, are satisfied with this monograph. What we have set out to do is: (a) to clarify the meaning and the basic purposes of evaluation; (b) to define and to explain the various steps involved in making evaluation *operational* in functional literacy programmes, from formulating evaluation policy to selecting indicators, designing instruments, collecting and storing data, and using such data to make informed decisions; and (c) to introduce, at an elementary level, the important concepts of evaluation design, instrumentation and statistical analysis as approaches basically towards making human impressions more objective, valuations more discriminating and conclusions more warranted.

Have we achieved these objectives? The Institute would welcome your verdict and would be pleased to share your opinions with the authors. *Write to: The Director, International Institute for Adult Literacy Methods, P.O. Box 1555, Tehran, Iran.*

To the memory of my father, Gurdial Singh Bhola

—lawyer, poet, psychic, critic, friend

What this book is about: Preface from the Author-editor

A preface usually introduces the subject and indicates the scope of a written work. It may also defend the choice of content and the way in which it is treated. Occasionally, therefore, the reader will be unable to follow the argument put forward in a preface unless he already has some degree of familiarity with the subject area involved. If the systematic study of educational evaluation is completely new to you, you may find it helpful to see how evaluation is defined in Chapter 1 of this monograph before reading this preface.

This monograph deals with the evaluation of functional literacy projects and programmes. Functional literacy, as commonly understood today, seeks to do more than teach the 3R's. It aims to help adults play a more effective role in their socio-economic and political milieu. Functional literacy programmes, therefore, are intended to have a major social impact on individuals, groups, institutions and communities.

The evaluation of functional literacy programmes raises problems common to the evaluation of other programmes with a similarly broad impact, such as adult education, lifelong education, non-formal education, community action and social development programmes. Nevertheless, while addressing the general interests of change agents in all such programmes, we are primarily concerned with the functional literacy worker. This monograph relates to his world of work, discusses his particular problems and employs his language. It should be useful to literacy workers in widely-differing ideological and cultural settings, whatever approach to literacy organization and teaching they may adopt and whatever the type of literacy programme with which they may be concerned.

Literature on the evaluation of functional literacy

Literature on the evaluation of education and development programmes has multiplied during the last ten years and includes material specifically written for literacy workers. Some of this, however, merely points out the need for literacy evaluation, sets out general guidelines for evaluators or presents individual case studies of project evaluation, occasionally accompanied by sample tests and questionnaires. At times, a single point of view is thrust upon the reader; at others, the various approaches adopted by evaluation experts are presented side by side, offering a bewildering choice of strategies to the beginner.

This monograph seeks to present a comprehensive and integrated approach to making evaluation *operational* within functional literacy programmes and projects. It will deal with all aspects of evaluation, from the design of evaluation policies to the making of needed tools and instruments, from data collection to the use of data to make informed decisions. This approach to evaluation will be referred to as the situation-specific strategy model.

The message—and the message behind the message

This monograph, *Evaluating functional literacy*, contains many explicit and implicit messages for the reader. Let us clarify some of these implicit messages by indicating what we wish to say and, perhaps more importantly, what we are *not* saying:

1. In writing a monograph on the evaluation of functional literacy, we are sending an implied message: we are advocating the need to evaluate functional literacy projects. However, we do *not* suggest that *all* literacy projects should be evaluated. In some cases, evaluation is quite unnecessary. For example, when a project is of limited scope and low risk, when its underlying assumptions have already been thoroughly tested, when its nature and format have previously proved successful, or when scarce resources must be put into action, rather than evaluation, the collection of a minimum set of management information may be quite sufficient.

2. We have, in this monograph, mentioned most of the current approaches to evaluation, including those variously labelled participative, naturalistic and bureaucratic. Discussion of participative and naturalistic approaches, however, is briefer than that of what we call the classical protocols of evaluation, *not* because we consider the former to be ideologically unacceptable or devoid of promise, but

because we feel that they have not yet developed the methodology needed to be operatively effective. Moreover, we wish to demonstrate that the situation-specific strategy model is inclusive of all other models and approaches. Evaluation policies can, after all, be designed in co-operation with clients; the choice of evaluation objectives and criteria can also be developed participatively. The situation-specific strategy model is not reserved for use by an independent outside élite working *on* clients rather than *with* them. One of its foremost strengths is that it can accommodate a wide variety of approaches and methods.

3. Why so much emphasis on design, measurement and statistics in this monograph? *All* approaches to evaluation have a goal; they seek a means of describing experience objectively in order to learn from it. To achieve this goal, evaluators need to master the basic concepts and techniques of design, measurement and statistics. These are the tools of the trade an evaluator uses to examine experience objectively in order to improve future actions.

4. Our situation-specific strategy model offers a complete and comprehensive approach to the evaluation of functional literacy. We suggest how the four parameters of the 'theory of general systems' (input, process, output and context) can be used to generate hundreds of evaluation questions. We also explain how to build elaborate systems capable of storing all kinds of data generated by the project or programme being implemented. However, while deliberately exposing all the possibilities, we are *not* saying that all or most literacy projects should conduct evaluation on so massive a scale. We are simply offering literacy workers an idealized system from which to select whatever they need in order to conduct evaluation as systematically as possible.

5. This monograph has been written particularly for use in training. We have, therefore, attempted to discuss with impartiality a wide variety of situations and approaches to evaluation. It is our hope that this monograph will prove equally useful to those already engaged in evaluation and those anxious to be so engaged, and that it will stimulate new applications and discoveries.

6. Lastly, we do not intend to suggest that evaluation is only for the experts—a special priestly class that no one else can join. However, we do suggest that evaluation is more than pooled ignorance. It requires expertise, but this expertise can be acquired. We, therefore, seek to demystify evaluation so that literacy workers at various levels can share in creating evaluative information.

The purpose of evaluation: making use of experience to improve future action

We suggest that evaluation is the process of using past experience to design better present and future actions. The task of the 'evaluation specialist', whatever approach he adopts, is to be systematic and objective in recording experience and deliberate in using that experience in future programmes.

This monograph sets out to help readers develop the skills needed for creating evaluative data and using these data to make informed programme decisions. Accordingly, we start by defining evaluation, and show it to be one link in the 'ideology to technology' chain. Next, the various evaluation models and approaches in vogue today are described briefly, emphasis is laid on the need to design evaluation policies appropriate to each particular programme in each particular situation and the situation-specific strategy model of evaluation is outlined. The process of making evaluation *operational* for functional literacy programme is then analysed. The measurement of change, and the collection and organization of data are discussed in detail, as are problems of statistical analysis and evaluation design, considered as means of drawing reliable and valid conclusions from the data generated.

The monograph is addressed to the middle-level evaluation worker in literacy projects and programmes. This person will not merely be making decisions *about* evaluation, nor will he be a field investigator setting out with questionnaire in hand to collect data for others. We have in mind an administrator or programme specialist who needs and can make use of objective information in making day-to-day planning, curriculum and organizational decisions within a project. He need not necessarily be officially designated an 'evaluator'; all programme specialists, teacher trainers, instructional materials specialists, field organizers and extension specialists who wish to apply their past and present experience to the planning of more appropriate future actions should find this monograph of interest. When an evaluation specialist is specifically assigned to a project team, other members must be able to understand his language, know what services he can provide and learn to collaborate with him without, unwittingly, becoming subservient to him.

If this monograph is intended for the middle-level literacy worker, why does it range from wide questions of policy to narrow problems of

data collection? Our desire to cover all aspects of the evaluation process springs from two considerations. First, policy formulators and programme planners in Ministries should know what evaluation is and how it works even though middle-level workers actually perform evaluation tasks. Secondly, the evaluator cannot perform his task satisfactorily unless he sees the whole picture, as he must be able to justify his choices and learn to collaborate with others in performing different evaluation tasks.

We have tried to make this introductory monograph as comprehensive as possible. Literacy workers in remote areas of the Third World, where there are often neither books to read nor peers with whom to discuss common concerns, should find within its covers a little of everything relating to their jobs. Understandably, the treatment of certain issues, questions and techniques may appear somewhat sketchy, but it is hoped that readers wishing for additional information on particular topics will eventually have access to more specialized materials.

The language used in the monograph has been kept as simple as possible. It is not, however, non-technical. Evaluators, as specialists, have a language of their own, and it is important to understand the terms they use in order to understand their ideas. Consequently, we have introduced those technical terms which we think a literacy worker, whatever his level and status, ought to know.

It is our hope that this monograph, while providing functional literacy workers with the skills and confidence they need to handle evaluation tasks, will at the same time help them to evaluate their own abilities and capabilities realistically. They should know what and how much they can do by themselves; and when, where and to whom they should turn for help if needed.

Finally, although I have written this monograph, many others have contributed to its development. Dr. John J. Hanitchak—at one time my colleague at Indiana University—and I started on this venture together, but changes in circumstances and interests extinguished the collaboration even as it began. David P. Rutt, Jr., my graduate assistant at Indiana University, rendered useful assistance. An earlier version of the monograph was presented to a 'Panel of Experts on the Production of Training Materials in the Field of Functional Literacy' which met in Dizine, Iran, during May 1976. The various members of that panel made excellent suggestions which I was able to incorporate in the revised version of the manuscript. I also had the good fortune to

be able to test this monograph in two test-in-use workshops on the evaluation of literacy programmes in Mwanza, United Republic of Tanzania (1976), and Mombasa, Kenya (1977), both sponsored jointly by the German Foundation for International Development and the International Institute for Adult Literacy Methods. These workshops led to further revisions. To these individuals and institutions I make grateful acknowledgement, even as I accept the responsibility for the inadequacies that still remain.

H. S. Bhola

Contents

		Page
	Invitation to the Reader	v
	What this book is about: Preface from the Author-editor	ix
Chapter	I: What is evaluation?	1
	II: The context of evaluation: evaluation as a link in the 'ideology to technology' chain	6
	III: Evaluation models and approaches	14
	IV: The situation-specific strategy model of evaluation	25
	V: Making evaluation operational in functional literacy programmes	34
	VI: Measurement of change: the classical protocols	54
	VII: Measurement of change: some rediscovered protocols	82
	VIII: Implementation data for management and evaluation	96
	IX: Special evaluation tasks for functional literacy workers	107
	X: Inspecting and analysing data: the use of statistics	121
	XI: Evaluation design: increasing the validity of conclusions	136
	XII: Interpreting evaluation results for decision-making	147

XIII: Concluding remarks 152

Appendix I: A selected annotated bibliography 157

II: Glossary of terms 161

CHAPTER ONE

What is evaluation?

To evaluate, according to the dictionary, is 'to examine and judge the worth, quality, significance, amount, degree or condition' of something.

Human beings are perpetually evaluating their past actions and experiences to maintain good performance, or to gain greater satisfaction by improving future performance. As individuals, we evaluate the significance of a Presidential speech for our business. We study the new lending rules of the credit co-operative with regard to the possibilities for future loans. We examine the quality of seeds we buy for next year's crop. We evaluate our personal interactions and social relationships to avoid emotional punishment or pain. Indeed, evaluation is part of the process of human survival.

Not only individuals but also communities and societies engage in evaluation. Unfortunately, neither individually nor collectively do we always make good evaluations. We are often tricked by our own limited perceptions. Sometimes we deliberately neglect to take note of the unpleasant and discomforting. Frequently, for political reasons, we declare our failures to be successes and defend our actions not with facts but with emotions.

Evaluation as a specialization

When we talk of evaluation as a specialization (as we shall do throughout this monograph), we are referring to a special body of ideas and skills developed during recent years. These ideas and skills can help us to draw accurate conclusions from our experience and to replace our subjective impressions of the world around us with objective findings that can be examined and shared.

This chapter opened with a commonsense definition of evaluation. Specialists, however, find such definitions hard to work with and prefer to develop their own conceptually-rooted ones.

Definitions of evaluation

We have already suggested that evaluation is intended *to examine experience objectively, systematically and with exactitude* in order to maintain performance at a certain level or to improve it.

Professor Lee J. Cronbach defines evaluation simply as 'the collection and use of information to make decisions about an educational programme'.

Dr. Marvin C. Alkin describes evaluation as 'the process of ascertaining the decision areas of concern, selecting appropriate information, and collecting and analysing information in order to report summary data useful to decision-makers in selecting among alternatives.'

To Professor Malcolm Provus, evaluation 'is the process of (a) agreeing upon programme standards, (b) determining whether a discrepancy exists between some aspects of the programme and the standards governing that aspect of the programme, and (c) using discrepancy information to identify the weaknesses of the programme.'

Professor Michael Scriven describes evaluation activity as consisting 'simply in the gathering and combining of performance data with a weighted set of critical scales to yield either comparative or numerical ratings, and in the justification of (a) the data-gathering instruments, (b) the weightings, and (c) the selection of criteria.' He then goes on to define two types of evaluation: formative and summative. Formative evaluation examines different aspects of the programme while it is being formed and is used to correct and improve the programme during its implementation. Summative evaluation sums up the impact of programme or project on client groups and communities.

Dr. Daniel L. Stufflebeam defines evaluation thus: 'Evaluation is the process of delineating, obtaining, and providing useful information for judging decision alternatives.'

Decision-making is the key phrase in all these definitions. Evaluation is useless unless the information it produces is used in decision-making. It must assist in clarifying decision areas and in establishing criteria for judgement and standards of performance. Information on performance must be obtained, analysed and reported to decision-makers to help them choose between decision alternatives.

So far, we have said what evaluation is. We should now explain what it is *not*, since the distinctions we shall make between evaluation and

research and between evaluation and supervision will prove useful at later stages of this monograph.

Evaluation is not research

Evaluation has often been confused with research, on the one hand, and with supervision, on the other. Confusion arises because evaluators collect numbers and scores, make tables and graphs and use tests and interviews, just as researchers do. However, although a carpenter may use the same tools to make a bridal bed or a coffin, the finished articles serve quite different purposes. The professional worker may play two roles—that of researcher and that of evaluator—and he may use the same tools: tests, questionnaires and interviews. Nevertheless, evaluation and research are, in functional terms, as different as the bridal bed and the coffin.

First, although the researcher and the evaluator may share a university background, the former's interests will probably be more academic and the latter's more action-oriented. The researcher normally chooses a topic that interests him personally, while the evaluator is briefed by a programme team and must be responsive to the needs of a group of decision-makers.

Secondly, the researcher may select an area full of theoretical significance, hoping to achieve a breakthrough in his field, but the evaluator must pick a subject of practical significance to an actual project and attempt to solve a specific problem. Thus, even when using the same tools, the researcher will be interested in perfect design, treatments without contamination and statistical significance; the evaluator will be pragmatic, using the best design possible under the circumstances, perhaps even inviting contamination in order to be realistic, and looking for practical, not statistical, significance.

Thirdly, the researcher may be at leisure to work at his own pace, but the evaluator is generally under pressure to produce some feedback within a given time limit, so that the programme team can make a new decision or confirm one already made.

Lastly, the researcher and the evaluator are themselves evaluated differently in terms of their performance. The former is judged according to whether he has made an original contribution to human knowledge and produced publishable research. The latter is considered successful if he has been able to provide the information needed by the decision-makers in a specific project to enable them to measure

outputs for the inputs applied.

A good evaluation study may serve immediate project needs and also contribute to human knowledge, thereby becoming what we have called research. Conversely, bad research may teach us less about human nature than a good evaluation study. Nevertheless, the distinction we have drawn between research and evaluation is a valid one and should be kept in mind.

Evaluation is not supervision

Evaluation, again, is *not* supervision; certainly not in the sense of policing subordinate officials within the organization or in the field. Evaluation, as we shall see, produces feedback on what is happening to the clients of a programme and what attitudes and skills they are learning. In functional literacy programmes, as in other programmes of innovation diffusion, planned change and development, it is usual to find a wide gap between intention and results. However, programme administrators often fail to analyse the reasons for this gap. Overlooking inappropriate models, communication barriers, and the divergent cultural experiences of the individuals and groups involved in the change process, they are apt to lay all blame on the field staff. This implies a rather naïve view of social and cultural dynamics, as discrepancies between intention and actual development are not always caused by staff irresponsibility. In any case the true objective of evaluation is to make analyses, not allegations, to find causes, not culprits.

The distinction between evaluation and supervision must be fully understood, not only by project administrators and planners but by *all* personnel. Fear should not drive subordinate staff to distort reality and falsify data in order to cover up their 'faults', thereby turning evaluation into a mere ritual and pushing planners into a world of fantasy that does not really exist.

Summary

Evaluation sets out to examine and judge the worth, quality and significance, amount, degree or condition of entities. As a specialization, evaluation can be defined as a process involving the collection and use of information to make decisions about educational program-

mes. It has developed a body of ideas and skills that can be useful in the collection and analysis of data. Important distinctions can be made between evaluation and research and between evaluation and supervision.

CHAPTER TWO

The context of evaluation: evaluation as a link in the 'ideology to technology' chain

Chapters I to IV of this monograph are devoted to the conceptual aspect, the 'what' and 'why', of literacy evaluation. Chapters V to XIII are concerned with operations and procedures demonstrating 'how' to evaluate functional literacy programmes.

It is easy to become impatient with abstract issues and methodological controversies when one is thirsting for action. If your interests are practical rather than theoretical, you may wish to turn to Chapter V immediately to find out how to plan, design and conduct evaluation studies. Nevertheless, you should read Chapters II to IV at some point, as a knowledge of evaluation policy and planning is a necessary part of the literacy evaluator's professional equipment.

Evaluation involves more than simply applying certain methods and techniques to a programme. It is a complex process that takes place within a socio-political context. Indeed, evaluation is only one link in the 'ideology to technology' chain shown in Figure 1.

The message in the diagram

Figure 1 presents a somewhat idealized situation, showing what should happen when a change is planned or its evaluation is undertaken. Under such conditions, evaluation would be the sequel to the change being planned. The change programme would evolve from national policies and plans adopted at country or community level, which would, in turn, emerge from the ideology of the nation. Change and evaluation methodologies would reflect the country's ideological and policy commitments, and the techniques and tactics chosen would also be consistent with the nation's political orientations.

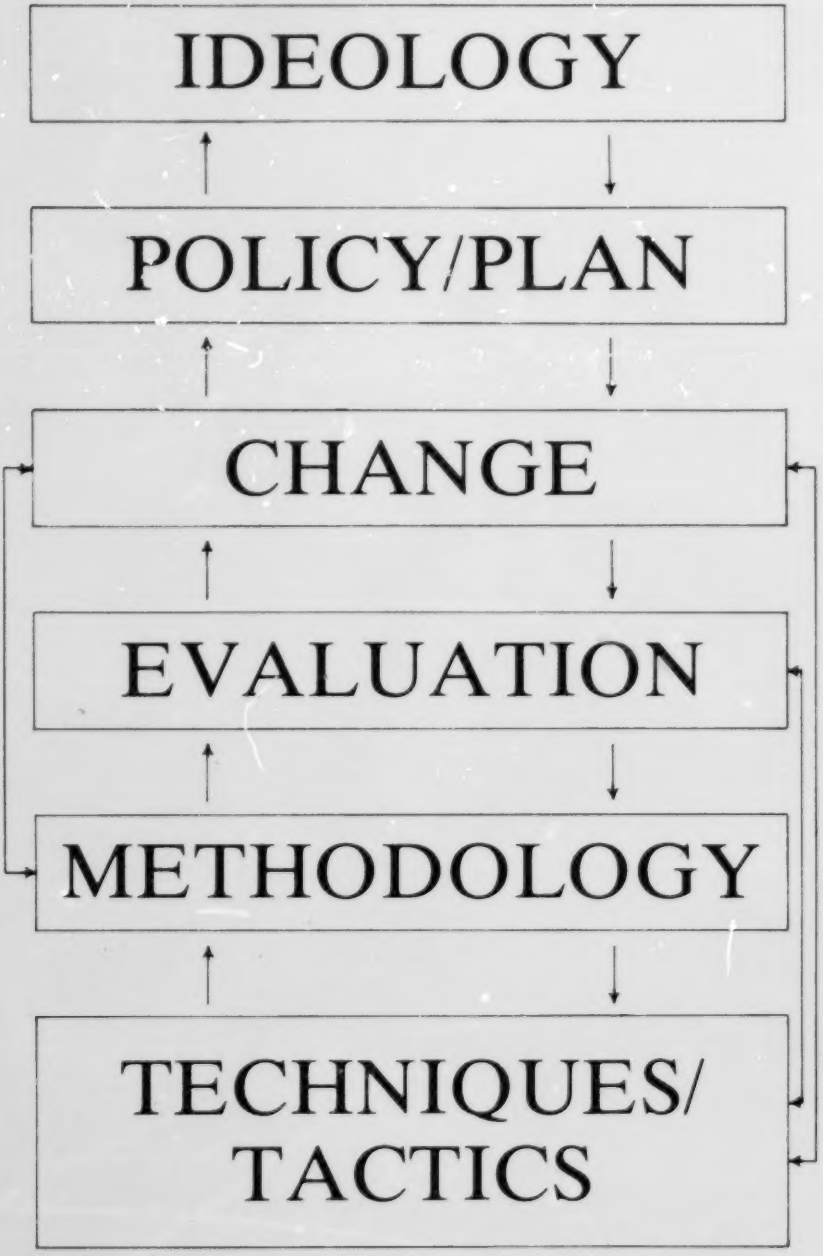


FIGURE 1. *Evaluation is a link in the 'ideology to technology' chain.*

The ideal, however, is seldom attained. The 'ideology to technology' chain gets tangled up, tied into knots and broken in places. The evaluator may rush to adopt a set of techniques without realizing that they are derived from conflicting methodologies. He may sometimes, methodologically speaking, be running with the hare and hunting with the hounds. While paying lip-service to humanism, he may in fact be using bureaucratic, even manipulative, techniques. Again, the criteria adopted in choosing methods and techniques of evaluation may conflict with those used in developing evaluation purposes and policies.

Lack of co-operation between those evaluating a change programme and other members of the programme team can also defeat the purpose of evaluation. Occasionally, although evaluators work hard, their limited understanding of the basic assumptions and objectives of the programme results in wasted effort.

Again, programmes of change and evaluation may fail to suit, or even oppose, national policies and plans. National policies themselves may prove ideologically barren or run counter to stated aspirations.

A good evaluator must be conversant with all the links in the 'ideology to technology' chain and be able to *articulate* them effectively. Let us now examine these links, and their implications for the literacy evaluator, in greater detail.

Ideology and evaluation

Ideology is visionary speculation. It is a set of ideas forming the basis of some economic or political theory, a view of the world that nations would like to create. Nations are inspired by their visions to plan better futures for their peoples. However, while all societies have ideologies, what is dear to the heart of one nation may seem worthless or despicable to another, and some nations do not act in accordance with their stated ideological aspirations.

In the life of a literacy evaluator, ideology is always around the corner. Ideological dilemmas will arise in the choice of methods and techniques, perhaps in terms of scientism versus humanism. Scientism is now felt to give too much emphasis to the methods of natural science, purely rational strategies and the dissection of human nature. It has come to be identified with behaviourism, bureaucratization and control. Humanism, on the other hand, refuses to reduce the individual to a mere object. The humanist talks of holistic methods,

participative approaches, mutual evaluations, non-manipulative techniques. His objective is growth, not control.

Scientism and humanism, however, are not necessarily diametrically opposed; philosophers even talk of scientific humanism. Nevertheless, the evaluator has to match his methods and techniques to the ideological climate of a country, although when collecting and organizing information there is no reason why the humanist should not count and measure, or use a prepared test.

Policy and evaluation

Policy and plans, which are concrete expressions of policy statements, form the link between ideology and the change programme. A nation's ideological aspirations must be stated in terms of national policies before they can be translated into action. Policy is a statement of intent by a ruling élite to bring about a new distribution of power in a society, involving a redistribution of status and of economic and educational goods. Politics can affect evaluation in many different ways and the evaluator must understand how and where the two intersect.

Evaluation as a political concern

This is the age of planned action. Design, not drift, is the new social ethos. We want to design our own destinies. We want to participate in the invention of our own futures. In most parts of the world today, more and more people are claiming the right to help shape policies and plans affecting their lives and work.

No wonder today's development programmes and projects are being subjected to close scrutiny. Their recipients want to know who is getting what, at what cost, at whose cost. They want to know whether the planners, designers, developers and change agents are doing the best possible job and making the right decisions. There is a political demand for educational, extension and development programmes to show tangible results. This concern with effectiveness and accountability can be met only through evaluation.

The competition for scarce resources

The growing interest in evaluation in developed and developing nations can be attributed, in addition to the increasing political

awareness of recipient communities and societies, to the competition for scarce resources. New applications of the social and behavioural sciences have produced many different strategies for use in the invention of new and better futures. Funding agencies and planning commissions, when faced with a variety of approaches to the fulfilment of the same objectives, naturally want to see comparative data on the outputs received for the inputs supplied.

In the area of functional literacy, literacy workers have to justify project expenditures. They are being asked: Is literacy a pre-condition for modernization? Can't the media carry the burdens of development communication? Isn't literacy merely delaying the process of development and consuming much-needed resources, which could perhaps be used differently with greater effect? When literacy projects *are* funded, literacy workers are being asked to demonstrate that they are doing a competent job and are actually fulfilling their stated objectives.

Apart from this genuine political need for evaluation, however, there are other somewhat more spurious ones. Donor governments and agencies often sell evaluation as part of the technical assistance package, even when the country concerned has neither the manpower nor the resources available to undertake it. In other cases, certain Third World institutions may be getting on the evaluation bandwagon simply to impress others.

The use and abuse of evaluation information

Evaluation produces information. Once available, information can be made to serve the selfish ends of various interests and groups. This is what is generally meant by the politics of evaluation.

The politics of evaluation can appear at all levels of decision-making. At one level, decisions may be made for or against the continuation, expansion or modification of a programme. If a programme is innovative, considerable time may elapse before it can make its full impact on a community; early evaluation may suggest that it is less effective than a more traditional one. *Formal* evaluation should, therefore, be avoided too early in the life of a programme, and the information it produces must be viewed in its true context. Again, where no systematic information is available, successes can be claimed where none exist, so that a programme that has been evaluated may compare unfavourably with one that has not been evaluated at all.

Comparisons must always be placed in perspective and fiction must be separated from fact.

At another level of decision-making, bureaucrats may make political use of evaluation to fight their private battles. Evaluation can reveal unfortunate discrepancies between intended results and actual impact. Thus, instead of providing feedback for use in programme improvement, it may become a stick with which to beat supervisors, an excuse for dismissing teachers and a means of branding competitors as dishonest and inefficient. If feedback is used correctly, it may raise awkward questions about goals, purposes, organizations, processes, strategies and deployment of personnel and resources. An evaluation study that raises such questions and creates the information needed to improve a programme may simply be withdrawn, kept secret and then quietly buried.

Evaluation and change

Evaluate is not an intransitive verb. You cannot just evaluate. You must evaluate something; that something is change.

The evaluator cannot do a good job unless he understands the change being planned and the model of change being used. He must know how the *means* and *ends* are related in the change plans. Are the means of change educational and attitudinal only? Are any material inputs being supplied? Are economic, social, legal and political structures also being changed or recreated so that new information and attitudes can be put to work? In other words, partial solutions must be separated from total solutions.

While a programme may be concerned solely with educational or attitudinal change, other related programmes may be intended to support it. In such a case, the change programme must be evaluated within the wider context of total programme actions.

Methodology of change and evaluation

Methodology is the science of methods. It is the principle and practice of orderly thought or procedure, arrived at by systematic analysis and application of the techniques of logic applied to a particular branch of learning.

In Chapter I, we made a distinction between research and evaluation. It is, therefore, reasonable to expect their methodologies to

differ. There is, of course, no one evaluation methodology. The evaluator may draw on several, including some normally associated with research, provided he ensures that those he chooses are ideologically compatible and embedded in the same world-view. In being eclectic, he must not become erratic.

The methodology of change may be structural, involving changes in political, legal or economic structures; it may involve material inputs; it may be informational and attitudinal. Again, the over-all approach to change may be revolutionary or reformist. In a change project with a reformist-educational orientation, the tactics of change may be either 'bureaucratic' (and, some might say, manipulative) or educational and growth-oriented.

Evaluation methodologies will also differ according to their basic views of man and of objective knowledge. In a change situation, the evaluator may prefer the collective perceptions of the people it will affect to a social scientist's report. What is to be evaluated? What standards should be used to measure success? How will 'data' be collected? What conclusions should be drawn from analyses and how? Such questions may be answered collectively; they need not be left to the trained social scientist.

Techniques and tactics

Technology can be defined in general terms as the application of science to the solution of practical problems. Once confined to the production of material things, the term is now employed in a wider sense, and can be used to describe the techniques and tactics developed to bring about improved learning or social change in a group or community.

The techniques and tactics chosen to bring about change must be consistent with the methodology adopted. Democratic participative approaches to decision-making at the community level cannot be taught using authoritarian methods. Humanistic approaches do not admit of propaganda, manipulation and behaviour modification. Similarly, if the evaluator claims to use a humanistic methodology, he cannot advocate standardized achievement tests for adult learners.

The choice of tactics employed to bring about change and the choice of instruments used in evaluation reveal ideological commitment or confusion only too clearly. The evaluator must clarify his position with regard to interpersonal relationships and objective truth. It is not

always possible to articulate the whole chain from ideology to technology, but the change-maker or evaluator must always attempt to do so.

Evaluation and ethics

The humanist may confidently assert that his whole approach to evaluation is an ethical one, but the scientifically-oriented evaluator will have to pay special attention to ethical considerations in the course of his work.

Evaluation often involves studying the effects of teaching and training people for new roles and settings. It entails questioning them about their income, family needs, sex habits, values, attitudes, hopes and fears. Such information once available can be used manipulatively, and even before it is used some important human problems arise. Individuals can be hurt by the questions put to them—even if they do not answer them. Evaluators may put undue pressure on their respondents in their eagerness to collect data and increase the percentages of returns, thereby encroaching on individuals' rights.

The use of information that does become available raises another string of questions with ethical import. Should it remain in the hands of the bureaucrats in charge of the programme, or should it be shared with the community that provided it? How should criteria for judging its usefulness be developed? How should evaluative data influence judgements and at what level should respondents participate?

Summary

In this chapter, we have placed evaluation in its total socio-political context. It has been presented as a link in the complex 'ideology to technology' chain. All the links in this chain have been defined and their inter-relationships have been clarified. The importance for the evaluator of understanding the nature of this chain and the ways in which its various links will affect his work is stressed. Finally, some ethical aspects of evaluation are discussed.

CHAPTER THREE

Evaluation models and approaches

Evaluation models have been developed by specialists in order to communicate the results of their own experience to others. They codify the general principles derived from years of evaluating educational programmes and projects. Some evaluators find the word 'model' somewhat pretentious and prefer to use the term 'approach'.

Why are there so many models of evaluation?

Why are we talking of evaluation models and approaches? Why has no one model or approach become universally accepted? The answer to these questions lies in the nature of the social sciences.

The social sciences study man and society. Each man holds his own opinions; each society has its own world-view. There is no universal agreement about human goals or the means of achieving them. As we pointed out in Chapter II, national ideologies differ and each one imposes a particular strategy of change and a particular methodology of evaluation.

Must the practitioner know all these models?

The evaluation practitioner needs to become familiar with the various models of evaluation for two reasons:

(i) A literacy worker engaged in evaluation must know which models exist if he is to select the one most suitable for his purposes in a given situation;

(ii) The field-worker must understand and speak the language of the expert if he wishes to engage in meaningful dialogue with him, safeguard his independence and be able to defend his choice of evaluation model.

In the following brief presentation of various evaluation models we

shall attempt not only to identify the basic concepts of each one, but also to see what it can offer the functional literacy worker engaged in the evaluation of a programme.

The CIPP model of evaluation

Daniel L. Stufflebeam and Egon G. Guba consider that the purpose of evaluation is to provide relevant information to decision-makers. The CIPP model suggests four types of evaluation:

(i) Context evaluation—the study of the setting of the change programme, undertaken to make planning decisions such as the setting of objectives;

(ii) Input evaluation—the evaluation of inputs made in the change programme, undertaken to make programming decisions, i.e. to develop strategies for reaching objectives;

(iii) Process evaluation—undertaken to implement decisions; and

(iv) Product evaluation—the evaluation of impact (the achievement of objectives), undertaken to make recycling decisions.

This model adopts the criteria of internal and external validity, reliability, objectivity, relevance, importance, scope, credibility, timeliness, pervasiveness and efficiency (see Glossary for the meanings of some of these technical terms). Evaluation must produce reports that decision-makers can use.

Although the CIPP model asserts that experimental design is not usable in conducting evaluations, it comes closest to formal *research* models. It stresses the clarification of goals and objectives and advocates structured observation as a means of discovering whether these goals and objectives have been achieved or not. It evokes the tradition of testing student achievement in schools and lays little emphasis on value concerns during the implementation or evaluation of a change programme.

WHAT CAN WE LEARN FROM THE CIPP EVALUATION MODEL?

Two lessons emerge from this brief description of the CIPP model:

(i) The scope of evaluation can be very wide. Context, inputs, processes and products can all be evaluated in relation to a particular change programme;

(ii) The sole purpose of evaluation is to produce information that can be used by the decision-makers. If the information produced is not usable, or is not used, evaluation must be considered to have failed.

The CIPP model also indicates the need to evaluate evaluation. The criteria of relevance, timeliness and credibility are particularly important. Evaluation must ignore irrelevant issues; it must produce information in time for the decision-makers to use it, and both evaluator and the information he produces must be trusted by the decision-makers. Lastly, the CIPP model asks us to look for non-experimental evaluation designs.

The discrepancy evaluation model

This model was proposed by Professor Malcolm Provus, who defines evaluation as the art of describing a discrepancy between expectation and performance of a programme.

Five different types of discrepancy can be located and studied in programme implementation. They may relate to: (a) the design of the programme; (b) its installation; (c) the process used to implement it; (d) its product; and (e) its cost.

According to this model, evaluation must be conducted by a team of evaluators who must reach agreement with the programme team on expected standards. Thereafter, every evaluation question must be answered by creating new information, judging performance against pre-established standards according to agreed criteria and making appropriate decisions.

This model shows a preference for the descriptive methods of history and anthropology and the case-study methods of sociology and psychiatry, but it recognizes the usefulness of the experimental design in certain cases.

It is claimed that this model can be used to provide continuous feedback and a direct one-to-one relationship between evaluative information and decision. However, full use of the discrepancy evaluation model would call for an abundance of experts, time and resources.

WHAT CAN WE LEARN FROM THE DISCREPANCY EVALUATION MODEL?

Once again, we see that the scope of evaluation can be very wide and that its main purpose is to feed back information to the decision-maker. The need for the evaluation and programme teams to collaborate in setting standards is deservedly stressed, and the model's suggestion that evaluation should be undertaken by a team indicates that more than one individual perception is needed if experience is to be described objectively.

Goal-free evaluation and similar approaches

Professor Michael Scriven originally defined evaluation as a methodological activity essentially concerned with comparing performance with goals. He recommended a holistic approach, using scientific methodology. Measuring scales were to be carefully designed, data on different performance scales were to be combined into a single rating, and both the weighting of scales and the choice of instruments used were to be justified. The selection of goals was also to be evaluated. A distinction was made between *formative evaluation* (for the implementers) and *summative evaluation* (for the consumers). The purpose of formative evaluation was to evaluate the processes involved in forming the programme—organizational strategies, curriculum development, motivational techniques, instructional materials, etc. The purpose of summative evaluation was to sum up the effect of the whole programme and discover what impact it had made.

In his later work, Professor Scriven proposed goal-free evaluation, concerned with the *actual* effects of a programme rather than its *intended* effects. He suggests that in goal-based evaluation, where the evaluator starts by examining programme objectives to see whether they have been attained or not, the evaluator's vision is so restricted that he may dismiss important consequences of a programme as mere side effects.

Recently, Professor Scriven has put forward the *modus operandi* method as a means of studying the cause-effect relationships. This method reconstructs 'the procedure of the historian, the detective, the anthropologist and the engineering trouble-shooter'. It is proposed as a substitute for experimental or quasi-experimental approaches when field situations preclude their use or to supplement them in order to increase the reliability of conclusions reached. Its methodology, however, is not yet fully developed.

WHAT CAN WE LEARN FROM THE GOAL-FREE EVALUATION APPROACH?

The lesson to be learned here is that goal-free evaluation methods should be included in all summative evaluations of functional literacy, as change programmes can bring about unforeseen—and sometimes undesirable—consequences, which may not come to light in an evaluation study based on a model tied to objectives.

Professor Scriven recommends holistic approaches, allowing both anticipated and unanticipated results of an action to be studied. He

favours the scientific method, urging the use of experimental and quasi-experimental approaches whenever possible, but recognizes the value of other methodologies, provided they are used with scientific rigour. It is worth remembering that the use of more than one methodological approach is a valuable means of ensuring accuracy.

The transactional evaluation approach

Although a number of evaluation specialists claim to use this approach, no one definition has yet proved acceptable to them all.

Professor Robert M. Rippey views transactional evaluation as a type of educational accountability whereby educators are asked to show what they have done rather than what scores their students have achieved. Transactional evaluation studies the change-makers themselves, examining their roles, the system in which they played those roles and the systems surrounding the system under change.

Professor Rippey distinguishes transactional evaluation from traditional formative and summative evaluation as follows:

> A comparison with traditional summative and formative evaluations shows that the target of evaluation is different: the subject of evaluation is the system, not the client or the services rendered by the system. The variables relate to the social, psychological and communication aspects of the system, rather than to the manifest objectives. The information is continuously fed back into the system. The evaluator himself is more a part of the operating system. The conventional considerations of reliability, validity and objectivity are less important than those of timeliness, relevance and the observable effects of generating evaluation information. Primarily, evaluation is intended to transform the conflict energy of change into productive activity; to clarify the roles of those persons involved in the programme changes, not to produce new knowledge or ascribe causality.[1]

WHAT CAN WE LEARN FROM THE TRANSACTIONAL EVALUATION APPROACH?

The purpose of evaluation is still to produce information, but the transactional evaluation approach is the first to direct attention to the change-makers themselves rather than to instructional materials and programme results.

This approach is highly value-laden. It seeks to produce psychological and relational information, not generalizable knowledge, explanations of causalities or proof of the superiority of a particular system. It

1. Rippey, Robert M., ed. *Studies in transactional evaluation*. Berkeley, California, McCutchan, 1973, p. 3–4.

seeks to convert the energy generated by conflict within systems into productive activity.

The transactional evaluation approach implies that evaluation should be conducted collectively by the protagonists and designers of a change programme and the representatives of those likely to be affected by it. It also urges sensitivity to the unanticipated consequences of change.

'Evaluation as illumination'

Some evaluation specialists have objected to the 'agricultural-botany' model used in the evaluation of so many educational change projects. This model, they suggest, gives adults an educational treatment (just as plants are given fertilizers or sprayed with pesticides) and then tries to estimate its impact (just as we measure plant growth). Their objection is that measurement of inputs and outputs may lead to neglect of the dynamics of change and that concern with quantity and numbers will, at best, result in only partial descriptions of phenomena.

Parlett and Hamilton[1] have proposed an approach, which they call 'evaluation as illumination', based on two important considerations:

(i) Instructional systems (formal education, functional literacy, community development, etc.), once adopted, become living systems. They undergo far from trivial modifications and eventually no longer fit their catalogue descriptions;

(ii) Programmes of instruction and development cannot be separated from the learning milieu. The actors in and structures of the learning milieu become part of the instructional system.

This approach to evaluation is a general strategy rather than a methodological package. It is prepared to retain certain features of the classical experimental approach: sampling techniques, the use of random samples, good questionnaires, even paper and pencil tests. Nevertheless, it is biased towards description and interpretation rather than measurement and prediction.

This type of evaluation will take place in three stages: (a) observation of the educational setting; (b) selection of themes through progressive focusing and intensive inquiry; and (c) analysis and explanation.

1. Parlett, M. and Hamilton, D. *Evaluation as illumination: a new approach to the study of innovatory programs.* University of Edinburgh, Centre for Research in the Educational Sciences, 1972. (Occasional paper, 9.)

A variety of methods, including observation, interview, questionnaires, tests, study of documentary and background materials, may be used. The approach can be applied to small- and large-scale projects and to the concerns not only of evaluation specialists but also of sponsors, participants, field-workers and other interested parties.

Social anthropological methods could be used to study various aspects of functional literacy projects: the structures of adult motivations; how much the adults learned; how literacy affected their personal and social lives; how much of what they learned they were able to use at home, at work or in the fields. They might also assist in discovering what is happening to the lives of teachers working in the field and in examining the inner life of the organization providing functional literacy services.

WHAT CAN WE LEARN FROM THE 'EVALUATION AS ILLUMINATION' APPROACH?

This approach, like the goal-free evaluation model presented above, rejects 'evaluation by objectives'. It emphasizes the dynamics of processes, recognizing that programme objectives change as programmes become living systems, and focuses on the milieu of instruction and change.

'Evaluation as illumination' accepts whatever methodological techniques will work in a given situation, but encourages the identification and analysis of themes by anthropological methods. Lastly, it points out that evaluation must serve the purposes not only of sponsors and administrators but also of field-workers and client groups.

The advocacy model of evaluation

The advocacy model of evaluation is evaluation in the legal mode. It is an educational trial by jury.

The proponents of this model (among them Professor Robert Wolf of Indiana University) suggest that the traditional approaches to educational evaluation, relying on achievement scores, interviews and survey data, are inadequate, since they provide only the correlates of instruction and extension rather than insights into the processes of change. Furthermore, evaluation is performed by experts for administrators and the views of programme recipients are not solicited.

The advocacy model of evaluation, in the elegant tradition of the common law, provides opportunities for all those involved in or

affected by a change programme to testify, regardless of their age, sex or status. Opposing views on the usefulness of the programme or project before the jury are put forward by a panel of educators and members of the target community respectively. The advocates of continuing support for the programme or project stress its positive aspects, while those wishing to transform or discontinue it present its negative ones. The jury then examine the logic of the arguments put forward and make an evaluation based on all the evidence presented.

The organization of an educational hearing involves four stages:

1. *The issue-generating stage.* The broad range of issues relating to the programme's objectives, organization, strategies and impact is formulated collectively by adults, field-workers, community leaders, specialists and administrators;
2. *The issue-selection stage.* The issues are reviewed by a representative panel, ranked in terms of their significance in the decision-making process and a suitable number are selected for the hearing;
3. *The preparation of arguments.* Each of the two teams of advocates prepares its case, referring to documents, minutes of meetings and test data and taking depositions from potential witnesses;
4. *The hearing.* The formal hearing may be preceded by exploratory sessions. Rules of evidence and procedure are established and cross-examination methods reviewed.

The advocacy model of evaluation sets out to present both sides of an issue in a balanced fashion, but most of its advantages are counterbalanced by corresponding disadvantages. The attempt to reflect different interests tends to give evaluation political overtones, it is difficult to find competent and committed people to serve on the jury and the role of the prosecutor is not easy. Nevertheless, the format allows all those affected by a change programme to voice their opinions and forces administrators to examine the possibility of discontinuing or drastically transforming programmes. When decisions will actually be made on the basis of evidence and arguments presented before the jury and when both teams of advocates are genuinely committed to the positions they are defending, the approach can produce significant results.

The following questions would lend themselves very well to the advocacy model of educational evaluation: Should literacy have been a component of an agricultural production programme for farmers and should the programme have been run throughout the year, or should it

have been suspended during the harvest season? What has been the impact of a functional literacy programme and was this impact worth the investments made?

WHAT CAN WE LEARN FROM THE ADVOCACY MODEL OF EVALUATION?

The principal message of this model is that the recipients of a change programme must be allowed to participate in evaluation by generating information about the programme, developing standards and criteria and judging. They must also participate in making decisions based on the conclusions drawn from the evaluative information generated. In other words, the monopoly of the evaluation specialist must be broken.

The methodological approach proposed is a holistic one, permitting the use of all kinds of data, from numerical to testimonial.

Participative evaluation

Paulo Freire, the author of *Pedagogy of the oppressed*, has promoted the participative approach to research and evaluation most successfully, although many others have spoken of it.

The methodology of participative evaluation is not yet fully formalized but its aim is to ensure that the people whose lives are being studied remain active participants rather than becoming mere objects. The participative approach assumes a particular concept of knowledge, does not use typical 'scientific' or 'objective' methods to find that knowledge and organizes evaluation in its own way.

For the participative evaluator, reality is not restricted to facts and physical objects, but includes the ways in which people involved with these facts and objects perceive them. It is thus the dialectical relationship between objectivity and subjectivity. The evaluator's task is to discover how people perceive themselves in their dialectical relationships with their objectivity. The evaluation design must make the connection between subjectivity and objectivity. The methods used must involve dialogue. The evaluator must involve the people being studied as evaluators; they must become both subjects and objects.

In an address to the Institute of Adult Education, University of Dar-es-Salaam, on 20 July 1972, Paulo Freire presented the possible steps in such a participative methodology:

1. The evaluation (or research) team should acquaint itself with all previous research and evaluation—no matter what methods were used in that research and evaluation.
2. The team should delimit the area of action geographically—even though, culturally speaking, there are no frontiers.
3. The team should identify official and popular institutions in the area selected and go to talk to the leaders within those institutions.
4. The evaluation team should tell these leaders, in all honesty, that they have come to discuss the possibility of *all* people in that community holding discussions and working together.
5. If the leaders agree, the evaluation team should hold meetings not only with the leaders of various institutions but also with the people who are involved in some way with those institutions.
6. The evaluation team should discuss with the community arrangements for meetings wherein groups of, say, thirty people could come together on a daily or weekly basis for discussions. Such meetings might involve almost all the inhabitants of a community and last for several weeks. The important thing would be to obtain a perception of the whole community.
7. Sociologists, psychologists, educators and linguists should, at this stage, join the research or evaluation team and visit each group. Records of discussions should be made at each meeting. People should be urged to speak if they are silent, but otherwise the role of the evaluation team should be no more than advisory. One of the members of the community should chair such meetings.
8. Justice, education, government, industry and many other topics may be discussed; but all in terms of the people and in the context of concrete realities.
9. When the smaller groups think they have exhausted the topics for discussion, each one should put its findings on paper and then they should all meet in a general session. The reporters at such sessions should be the people themselves; not the specialists on the team. The workers should become intellectuals. There should be collective discussion of each group report.
10. The evaluation team should now make a critical study of the people's discourse. This study should be interdisciplinary. The various levels at which people perceive reality must be determined and their many implications should be worked out. These implications must be studied in the presence of the people, not by social scientists on their own.

11. The evaluation team together with the people should now draft a proposal for subsequent action. The programme itself should not be worked out *for* the people but *with* the people.

Thus, participative research and evaluation are not research and evaluation as commonly understood. They are not distinguishable from pedagogy or need assessments. They educate and raise the consciousness of people. The participative approach is a prescription for action by human beings working together; it does not seek to record to objectify and to evaluate; it seeks to transform reality in the very process of defining it.

WHAT CAN WE LEARN FROM THE PARTICIPATIVE EVALUATION APPROACH?

The participative evaluation approach teaches us humanism. It reminds us that the whole purpose of education and change is to touch the lives of individual men and women. It warns us not to manipulate them or make them dependent on an authority structure. We must enable them to grow and become responsible for changing the realities that surround them.

This approach abolishes the distinction between instruction and evaluation, and is certainly useful at the community level or within group settings. However, it fails to generate all the different types of data that might be needed by administrators and decision-makers, and should, therefore, be employed in conjunction with other forms of evaluation.

Summary

In this chapter we have talked of various recently-proposed models and approaches to evaluation: the CIPP model of evaluation; discrepancy evaluation; goal-free evaluation and similar approaches; transactional evaluation; 'evaluation as illumination'; the advocacy model of evaluation; and participative evaluation. The lessons that a functional literacy worker can learn from each of these models have been underlined. In the following chapter, we shall show how some of these lessons might be put to use in literacy evaluation.

CHAPTER FOUR

The situation-specific strategy model of evaluation

The previous chapter outlined a variety of existing evaluation models and approaches. Some of these are devices for *ordering* the elements in the evaluation process into a harmonious system, by developing general orientations, defining different kinds of decisions and the evaluative information needs of decision-makers, or identifying discrepancies between programme objectives and performance. Other models are methodological, either oriented towards design and quantification or rejecting classical research models in favour of naturalistic methods. Some approaches advocate the goal-free exploration of effects, others propose participative formats for the use of data in evaluation and decision-making, and at least one suggests the abolition of the distinction between evaluation and instruction, considering evaluation to be part of social *praxis*.

All of these models are good, but none is totally comprehensive. Some are obviously culture-bound; others are highly ideological. A single model cannot always answer all evaluation questions, as even the participative methods rely on the use of data, which may have to be produced by tests and questionnaires or even specially generated. Some approaches have not yet been fully developed or tested in the field. The models described in the previous chapter, therefore, as even their proponents admit, should be used not in isolation but to complement other more traditional models of evaluation.

The situation-specific strategy (3-S) model of evaluation

In our earlier discussion of evaluation as a link in the 'ideology to technology' chain, we demonstrated that each ideology will suggest its own social change policy and methodology, which in turn will evolve particular methodologies and techniques of evaluation. These methodologies and techniques will vary according to the particular

33

situation in which evaluation is undertaken; they will have to be situation-specific. Every programme must define its own approach to evaluation and design an evaluation methodology to suit its own needs and objectives. The 3-S model outlined below can be adapted to all ideological, social and technical requirements and is consequently able to provide the most appropriate evaluation strategy for any change programme.

This model not only enables us to conduct evaluation in a variety of social and ideological contexts, but also allows us to draw upon a variety of methodologies, and it is our belief that the new evaluation techniques proposed in the previous chapter deserve to be studied but cannot be used effectively unless the evaluator is already thoroughly familiar with the basic concepts of research design, measurement, theory and statistics, which offer well-tried techniques for sampling experiences, interpreting quantities and making deductions.

What is our position?

In proposing a 3-S model of evaluation that accommodates plural frameworks and multiple methodologies, what position are we adopting? In trying to defend everything, are we in fact defending nothing? Are we being conceptually evasive?

We do not think so. Far from condoning incoherence, we recognize that the evaluator must possess intellectual integrity and independence. He must not be influenced by 'group think', the psychosis of a particular professional community or the empty slogans of society. If the emperor has no clothes, the evaluator should be strong enough to say so. The good evaluator must make it possible for the *technical process* to interact with the *social process* by offering the best available technical options for generating the required information and by negotiating his values and choices with all those affected by the programme system.

We are not unaware of the differences between models and methodologies. We do wish to show, however, that although models may vary as to their ideological orientation, they may share a similar preference for the study of processes rather than outcomes, the use of holistic methods and the participation of client groups. In many cases, therefore, amalgamation of models and approaches is both possible and fruitful. Our 3-S model enables evaluators to examine the full range of models and approaches before choosing the strategy best

suited to the situation in which evaluation will be undertaken. We are offering an 'à la carte' menu and not a set dinner.

Developing a 3-S model of evaluation

The development of an evaluation strategy based on the 3-S model should comprise the following five steps: (a) ordering the world of change and evaluation; (b) articulating the means-ends relationships in the change programme; (c) generating profiles of information needs and evaluation issues; (d) developing a situation-specific evaluation agenda; and (e) choosing appropriate and realistic methodologies and techniques.

ORDERING THE WORLD OF CHANGE AND EVALUATION

Ordering the world of change and evaluation involves finding out what is going on. What is the change in question? Who is changing whom? Who are the evaluators? What are they evaluating? How are they related to the decision-makers?

We must start by using our 'social eye'. We must identify the communities involved in the change and evaluation processes and the institutions, both within the community and outside it, engaged in the change process. We must look for groups within and outside these institutions and examine possible overlapping. We must find the individuals who, on account of their influence or formal positions, can play strategic roles in the change and evaluation processes.

A systemic view of change and evaluation

In ordering the world of change and evaluation, we will find it useful to consider two related concepts: system and network. What is a system? A system, simply, is a whole with interdependent parts. If something happens to one part of a system, its other parts will also be affected in some way. The human body is a system; if something happens to one part of it, the whole body will be affected. Furthermore, the human body is a system made up of overlapping subsystems: optical, nervous, respiratory, digestive, etc.

The system metaphor is useful because the change system, like any other system, can be described in terms of four parameters: context, input, process and output.

Context is what surrounds the system—its environment. Using our

social eye, we identify the limits of a system of change, separating it from its environment by imaginary system boundaries, but remembering that the environment is ever-present and remains a potential influence on our change efforts within the system we have outlined.

Input is what is put into a programme system—men, materials, tools, etc. This parameter helps us to think about what is going into the system to bring about the change.

Process deals with the 'how' of procedures, arrangements, formations, extension and education. When dealing with systems under change, we may have to consider a number of processes: informational, organizational, distributive, etc. The process parameter helps us to take stock of the processes brought into play within the system of change.

Lastly, *output* is what is put out at the end—the product of the system. What is the system producing? Is it learning? Is it increased agricultural productivity? Is it trained manpower? Is it the new socialist man?

This systemic view of a change programme should help us to order the world of change. What individuals, groups, institutions and communities form the programme system? Is there a system of action extending from the nation's capital to the field? How is this system of action connected with the client system? Are the change-makers in direct contact with the ultimate clients, or are they separated by a chain of intermediaries?

The system metaphor can be applied to evaluation as well as to change.

Context evaluation would be concerned with analysing, defining and measuring the environment of the programme system. It would be helpful in planning decisions. (What kind of climate for change may be experienced in a particular situation?) It could predict results in addition to measuring impacts.

Input evaluation would be concerned with determining alternatives in the utilization of resources to achieve programme goals. It would also suggest what regions, occupational groups or economic sectors might receive priority for change and why.

Process evaluation would be concerned with designing procedural strategies and comparing the effectiveness of different approaches to instruction, extension, animation and organization.

Finally, *output evaluation* would be concerned with determining the quantity and quality of the programme product and the existence of

any unintended social or economic by-products.

The system metaphor can be applied to the evaluation system in yet another way. The context of the evaluation system itself can be defined. (Is the environment sympathetic to the evaluation effort?) The inputs into the evaluation system can be weighed. (Do we have enough evaluation personnel? Are they sufficiently trained? Do material resources exist to finance the evaluation effort?) Similarly, questions can be raised about the process and product of the evaluation system. (Is the evaluation process sufficiently objective? Is the product of evaluation merely a report or does it improve the programme staff's understanding and, thereby, improve the programme of action?)

Identifying change and evaluation networks

A network is an interconnected or interrelated chain; in other words, a system. The word network is used here to represent a group of individuals linked together in chains of command or co-operation. In ordering the world of change and evaluation one has to identify the network of decision-makers in the change system and the network of evaluators in the evaluation system and then discover whether these two networks are themselves interfaced, that is, interconnected.

It is obvious that in most planned social change we are dealing not merely with an individual change-maker or group of change-makers but with a whole system of action. This system may extend from the field worker to a key Minister or the Chairman of a National Planning Commission. The change network and the place in it of every individual engaged in change must be identified so that each one sees his work, responsibilities and prospects in perspective.

Similarly, the evaluation network must be identified. What persons constitute the evaluation network? Are they also the decision-makers in the programme system? In other words, are they playing dual roles—that of change-maker and that of evaluator? Is there a mixture of independent roles and dual roles in the evaluation network? How far does the evaluation network extend? Is it properly interfaced with the programme system at all levels so that information generated or collected at each level of the evaluation system can be used by the decision-makers at the corresponding levels of the decision-making network?

Internal and external evaluation

Evaluators must also consider the division of labour with regard to the evaluation of a programme. Should evaluation be carried out by insiders—programme specialists and evaluators specially assigned to a project? Or should it be carried out by outsiders in no way related to the project's design or implementation? These questions are bound up with another raised earlier in this monograph: is the purpose of evaluation to exercise *control* or to encourage *growth*? In practice, evaluation has to serve both these masters: growth *and* control. Accordingly, we suggest that process-related evaluation for programme improvement might be conducted by insiders so that they learn from the process of evaluation itself, while evaluation of the programme's impact on communities and institutions could be studied by outsiders. These issues must be anticipated and, if possible, resolved in the process of ordering and network identification.

The insider-outsider question raises another issue. When evaluation is handled by insiders, should it be entrusted to one evaluator and his associates or to the whole project team, including its programme specialists? We wish to assert that evaluation should not be monopolized by evaluators. It should be treated as a group task at every stage, from that of planning to that of generating usable information from the data collected. Not only should evaluation be planned by the project team, it should also be conducted by the whole group, with each member making his own special contribution to the collective effort. If evaluation is, as we have defined it, a process of servicing decisions—that is, of serving decision-makers—how can it take place if the decision-makers are not consulted? Unless the whole team sits down to analyse the programme, define parameters, list the variables involved, anticipate likely decision settings and articulate the information needs for making those decisions, how can the evaluator do what he is supposed to do? At this early stage of programme analysis, the evaluator should listen and ask questions but should not say what is worth evaluating and what is not. That initiative must lie with the programme specialists. If these specialists do not take the initiative at this stage, they probably understand neither the programme nor its structure and will consequently be unable to implement it, let alone assist in the development of an evaluation system.

In devising an evaluation methodology and preparing questionnaires, observation inventories and interview schedules, the evaluator

will play a key role because he is technically trained to perform those tasks. However, data collection should *not* devolve on him alone, since it concerns the whole team. Field workers, supervisors and area administrators can collect data while carrying out their other duties; a parallel bureaucracy of evaluators should not be created merely for data collection. Information created by evaluation must be *used* by programme implementers, who will obviously have more confidence in information they themselves have helped to create and will develop professionally by participating in evaluation.

ARTICULATING THE MEANS-ENDS RELATIONSHIPS IN THE CHANGE PROGRAMME

Once the change and evaluation systems have been ordered and the existing networks of decision-makers and evaluators identified, it is time to take the next step. The means-ends relationships implicit in the change programme must now be articulated. In other words, the focus should now shift to the *strategy* employed to bring about change.

Some change strategies may be purely informational and educational, assuming that once people have been given new information, or taught new attitudes and skills, they will know how to use this knowledge. Unfortunately, however, the provision of education alone does not always ensure change. Other change strategies may involve material inputs as well. Information about new health attitudes may be accompanied by the provision of village clinics and medicines. When new methods of farming are taught, new seeds and agricultural credit may also be supplied. Finally, some change strategies may also involve structural changes. New legislation may be passed to redistribute lands, land consolidation and marketing co-operatives may be created and farm insurance may be made available. In other words, the rules of the political and economic game in a society may be changed.

The evaluator must understand the strategy being used in the change programme with which he is concerned. It is unlikely that informational, material and structural strategies will all be used within the same programme. There have been very few multi-strategy programmes. Nevertheless, a single-strategy programme often assumes that another programme in the same community will use a complementary strategy. Thus, a literacy programme may expect to work alongside an agricultural extension programme in a change project for a single community. The evaluator must, therefore, understand the underlying assumptions and expectations. He should be able

to identify the individual contribution of each programme and to evaluate his particular programme in relation to the total change effort.

GENERATING PROFILES OF INFORMATION NEEDS AND EVALUATION ISSUES

In this step the evaluator and programme specialists must work together to bring about a confrontation of the change system and the change strategy so that the information needs of decision-makers can be defined. Everything the latter might wish to know should be listed. What kinds of information would be needed for the simple management of the change programme? Would the programme generate this information in the natural course of its implementation? Would special mechanisms have to be created to generate, collect and store such information? The evaluation questions that, if answered, would create needed knowledge and illuminate relationships between different variables must also be listed. Such a list, without becoming irresponsible, should be as comprehensive and theoretically complete as possible.

DEVELOPING A SITUATION-SPECIFIC EVALUATION AGENDA

The lists of information needs and evaluation questions should now be reviewed by programme specialists and evaluators to develop a situation-specific evaluation agenda. Evaluation costs money. When there is a choice between evaluation and the provision of services to the poor and underprivileged the answer should be obvious: the poor must win. Evaluation plans for a pilot programme must be conservative. Evaluation agendas should be realistic; nothing is more unfortunate than an evaluation agenda so ambitious that it could never be fulfilled or one that, when implemented, produces more data than can be used by the decision-makers.

CHOOSING APPROPRIATE AND REALISTIC METHODOLOGIES AND TECHNIQUES

Once the situation-specific information needs of a project or programme have been established and an evaluation agenda has been determined, methodologies and techniques, again situation-specific, must be chosen. These methodologies and techniques should be appropriate for the types of evaluation questions raised, as well as realistic in situational terms. If the services of a trained clinical psychologist are not available, it is best to leave in-depth analyses of personality changes in adults out of evaluation plans. If computer

analysis of data is not possible, one must be content with paper and pencil analysis and adjust inventories and questionnaires to the realities of the situation. In sum, evaluators must seek to rework experience as systematically and accurately as possible in the prevailing circumstances.

Summary

In this chapter we have presented the situation-specific strategy model of evaluation. The five stages in its development are: (a) ordering the world of evaluation and change by conceiving it in terms of systems and networks; (b) articulating the means-ends relationships implicit in the change programme to clarify the strategy of change; (c) generating profiles of information needs and evaluation issues through the interaction of the system and the strategy of change; (d) developing a situation-specific evaluation agenda for a particular programme at a particular place and time; and (e) choosing methodologies and techniques that are technically appropriate and situationally feasible.

CHAPTER FIVE

Making evaluation operational in functional literacy programmes

The first four chapters of this monograph attempted to conceptualize the process of evaluation. After defining evaluation, we established certain relationships between this and similar processes. We examined evaluation in the perspective of policy and planning and showed it to be one of the links in the long chain leading from ideology to technology. After drawing thumbnail sketches of a number of currently controversial models and approaches, we finally proposed and described an integrated evaluation model, the situation-specific strategy model.

From the conceptual, we must now move on to the operational. We are often frustrated by our inability to put attractive and apparently sound ideas into practice. In other words, we find it difficult to translate certain concepts and ideas into identifiable, concrete and repeatable operations and actions. The basic concern of this monograph is to explain how evaluation can be made *operational* for broad impact functional literacy or general development programmes. The evaluation model outlined above must now be put into action.

Making evaluation operational in a functional literacy programme entails the following three processes: (a) programme system articulation; (b) evaluation system definition; and (c) data system design.

One cannot begin to develop a programme, let alone a system of evaluation, without first analysing and articulating the means-ends design of the development programme (or programme system) it will support. Similarly, questions about what data should be collected and how that data should be stored cannot be raised unless the evaluation system has been defined. Before defining an evaluation system, a data system cannot be designed.[1]

1. Of course, not all data systems are established by evaluators. A data system can be planned independently of an evaluation system when it will *not* be used for evaluation, but merely as an accounting and auditing system.

Let us restate the argument more fully. To develop a functional data system that can be used to provide useful information,[1] one must collect and store data that are significant for programme specialists and decision-makers. When establishing a data system, we should not be interested in *any* data or *all* data on client groups, change agents, instructional materials, teaching settings and organizational processes; we should be interested only in significant data. But what are significant data? This question cannot be answered in a vacuum. We must ask: What information will the evaluator need if he is to perform the needed evaluation tasks? This brings us back to the evaluation system, proving that in order to build a data system, one must first know the objectives, scope and structure of the evaluation system being planned.

Evaluation is rightly defined as the process of obtaining, creating and providing information to enable programme sponsors, planners and implementers to choose between alternatives. It is thus the process of *servicing* decisions. But, we may ask: What decisions? What types of decisions? Decisions about what people, what patterns, what structures and what operations? These questions cannot be answered in a vacuum either. We must know the programme system if we are to understand what actions and, consequently, what decisions are involved. We will also have to decide what important decisions *must* be serviced by evaluation based on hard data. Evaluation and information collection require time and money. Very often, planners may want not an ideal evaluation system and an ideal data storage system but systems they can afford. This point raises questions of evaluation policy to which we will return later.

To recapitulate, a *data system* cannot be planned independently of the *evaluation system*. Similarly, an evaluation system cannot be planned independently of the *programme system*. Figure 2 depicts their interdependence in graphic form.

Programme system articulation

The process of making evaluation operational must start by examining the programme. In our case, for example, we must analyse the nature

1. A distinction between data and information is sometimes made by specialists in information science. They suggest that information is developed for use from stored data; data is what can be used to produce information for a decision-maker. In this monograph these terms are often interchangeable.

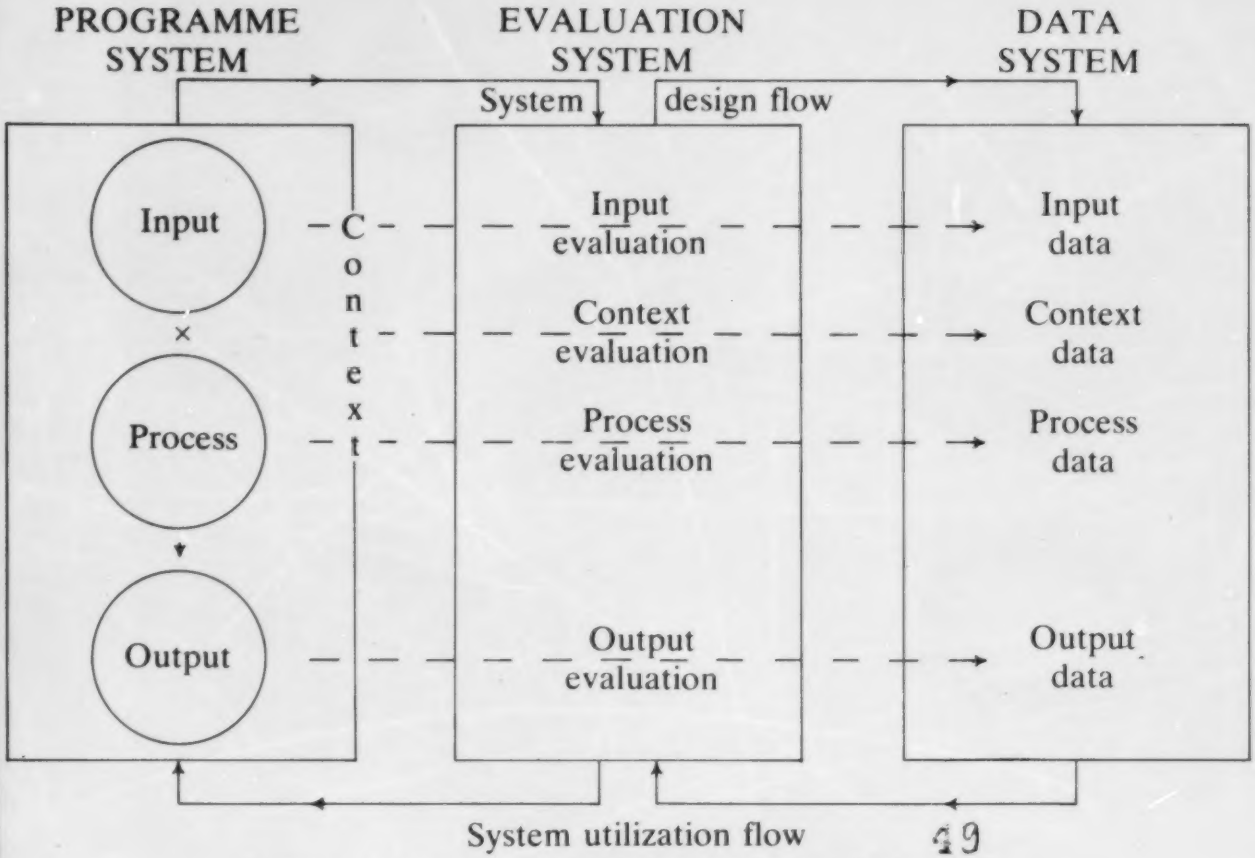


FIGURE 2. *Interdependence of programme, evaluation and data systems.*

of the functional literacy programme. What is the concept of functional literacy and what programme will result from implementing that concept? What do some functional literacy workers mean by asserting that the objective of such a programme should be to teach the world, not merely the word?

We must develop a generalized model of functional literacy comprising the four system parameters: context, input, process and output. This generalized model must be used to articulate the means and ends relationships specific to the situation. What inputs are actually being made? What processes are entailed? Where will impacts of the programme appear? Again, as part of the programme system articulation, the system of action (or network of decision-makers) must be identified. Questions must also be asked at this stage about the existence of other development programmes which might be considered to supplement or complement the functional literacy programme in bringing about change within a single community.

THE PROGRAMME: WHAT IS FUNCTIONAL LITERACY?

To be functional is to have a function. Nevertheless, although it would be somewhat absurd to think of literacy as anything other than functional, the term 'functional literacy' has been interpreted in a number of different ways. Unesco, in the context of its Experimental World Literacy Programme, defined it as follows:

> Briefly stated, the essential elements of the new approach to literacy are the following: (a) literacy programmes should be incorporated into and correlated with economic and social development plans; (b) the eradication of illiteracy should start within the categories of population which are highly motivated and which need literacy for their own and their country's benefit; (c) literacy programmes should preferably be linked with economic priorities and carried out in areas undergoing rapid economic expansion; (d) literacy programmes must impart not only reading and writing, but also professional and technical knowledge, thereby leading to a fuller participation of adults in economic and civic life; (e) literacy must be an integral part of the over-all education plan and educational system of each country; (f) the financial needs of functional literacy should be met out of various resources, public and private, as well as provided for in economic investments; (g) the literacy programmes of this new kind should aid in achieving main economic objectives, i.e., the increase in labour productivity, food production, industrialization, social and professional mobility, creation of new manpower, diversification of the economy.[1]

1. *An Asian model of educational development (perspectives for 1965–80)*. Unesco, Paris, 1968, p. 97.

There are important implications for the functional literacy worker in the above definition with regard to over-all policy planning, programme development, teaching methodology and choice of programme content. The Unesco concept of functional literacy has been very influential indeed. Most literacy workers in the world today insist upon teaching their adult learners economic and political skills in addition to reading. The Declaration of Persepolis, issued at the end of the International Symposium for Literacy, 3–8 September 1975, considered literacy to be:

> . . . not just the process of learning the skills of reading, writing and arithmetic, but a contribution to the liberation of man and to his full development. Thus conceived, literacy creates the conditions for the acquisition of a critical consciousness of the contradictions of society in which man lives and of its aims; it also stimulates initiative and his participation in the creation of projects capable of acting upon the world, of transforming it, and of defining the aims of an authentic human development.

Literacy is thus perceived as fundamental social change, an experiment in human liberation.

A GENERALIZED MODEL OF THE PROGRAMME SYSTEM

As we indicated earlier, any broad impact programme of social development and change can be seen and described as a *system*. A functional literacy programme can, therefore, be described by answering the following questions:

1. What inputs can probably be made into a functional literacy programme?
2. What operations and processes are likely to be used within the programme to convert inputs into required outputs?
3. What outputs are normally intended and what might be some unintended outputs?
4. What is the probable socio-cultural context or environment of such a programme that might interact with and affect inputs, processes and outputs?

We must now list, categorize and describe in general terms variables and treatments which might enter such a change programme. In order to do so, we must comprehend the change model on which a programme of functional literacy is based, and also the general objectives of such a programme. If we do not understand the concep-

tual structure of the programme model being used or the programme objectives, we may be listing inappropriate variables.

Let us first try to list the *inputs* in a functional literacy project. Since we are concerned at present with a generalized model, this list will be equally general. Any individual functional literacy project should be able to adapt this list to its own specific needs with a minimum of effort.

The inputs, we suggest, can be of three kinds: (A) Clients, (B) Change agents, and (C) Materials and facilities.

1. *INPUTS*

A. *Clients:* (a) Adults (men and women); (b) Families; (c) Community groups; (d) Traditional institutions and formal institutions (co-operatives, unions, etc.); and (e) Communities and sub-cultures[1].
B. *Change agents:* (a) Literacy teachers; (b) Supervisors; (c) Extension workers; (d) Forum leaders; (e) Administrators; (f) Technical specialists; and (g) Community leaders.
C. *Materials and facilities:* (a) Instructional materials; (b) Production equipment and supplies; (c) transportation; (d) Plants, spaces and housing; (e) Duplication and printing facilities; (f) Broadcasting facilities; and (g) Time.

Let us now describe the second dimension of the programme system—*operations* and *processes*. At least three categories of operations and processes may be associated with a functional literacy programme: (L) Organizational/structural, (M) Instructional/formative, and (N) Distributive/maintenance-related.[2]

2. *OPERATIONS/PROCESSES*

L. *Organizational/structural:* (a) Class organization; (b) Organization for discussion forums; (c) Organization for demonstration/extension work; (d) Organization for literacy/development committee; (e) Project organization; (f) Structuring expert-counterpart relationships;

1. Depending upon the unit of analysis and the level of response, communities and even sub-cultures could be considered as inputs. In some cases it might be more useful, however, to consider communities and sub-cultures as context variables, that is, as the environment of change.

2. Clearly, there are other processes which could conceivably be used in programmes of planned change and development. The objectives and structures of most functional literacy programmes would suggest, however, that these three categories are most typical, subject to design by functional literacy workers.

(g) Interdepartmental co-ordination, and (h) Institution-building and organization development.[1]

M. *Instructional/formative:* (a) Teaching, training and extension methodologies for (i) individuals, and (ii) groups; (b) Pre-service and in-service training strategies for staff; and (c) Methodologies regarding orientation of international experts and national officers from co-operating departments at various levels.

N. *Distributive/maintenance-related:* (a) Distribution of material and supplies; and (b) Transportation and circulation of personnel.

We may now categorize and list the variables under the *output* dimension of the system. We have indicated earlier that the inputs, processes, outputs and context of any programme system cannot be listed without knowledge of the conceptual model on which the system has been based and knowledge of the programme's objectives. The relationship between *outputs*, on the one hand, and *objectives*, on the other, is direct and should be clearly understood. The outputs of a functional literacy project may be categorized as follows:

3. *OUTPUTS*

P. *Individuals who are:* (a) Literate; (b) More productive; (c) Healthy; (d) Active participants in socio-political institutions; (e) Efficient consumers of information and education; (f) Holders of modern attitudes towards consumption and investment; OR any alternative set of individual characteristics.

Q. *Institutions that are:* (a) Secular/modern; (b) Functional; (c) Responsive; OR any alternative set of institutional characteristics.

R. *Communities that:* (a) Are cohesive; (b) Are healthy and attractive; (c) Provide a literate environment; OR any alternative set of community characteristics.

S. *Knowledge for professional literacy workers about:* (a) Preconditions for functional literacy to play a role in development; and (b) Design, implementation and impact of programmes.

The *context* variables of a system can be similarly analysed in order to study their impact on the variables of inputs, processes and outputs.

4. *CONTEXT/ENVIRONMENT*

X. *Political context:* (a) Supportive factors; (b) Neutral factors; and (c) Inhibitive factors.

1. This list makes implicit assumptions about multiple inputs; it can accommodate analyses at psychological, institutional and sociological levels and its perspective on reality is best described as configurational.

Y. *Economic context:* (a) Supportive factors; (b) Neutral factors; and (c) Inhibitive factors.
Z. *Social-cultural context:* (a) Supportive factors; (b) Neutral factors; and (c) Inhibitive factors.

This generalized model of functional literacy will obviously have to be adapted by individual projects to suit the situation-specific programme system with which they are working.

FROM THE GENERAL TO THE SPECIFIC: THE MEANS-ENDS RELATIONSHIPS

The system outlined above is, of course, highly idealized. It displays the full range of possible functional literacy inputs, processes, outputs and contexts. Every functional literacy programme, however, will have to specify the ends it seeks to achieve, and the means it will employ to achieve them. In other words, a means-ends calculus will have to be defined for each particular functional literacy programme in a particular geographical location at a particular historical time. The following kinds of question will have to be answered: What is the cultural and political context of this programme? Is the context of this programme so clearly defined that it will determine the inputs that can be made and the processes that can be employed? Or is the context neutral, allowing functional literacy workers a certain amount of freedom as to the choices that they can make? What kind of inputs are available in this particular programme and how flexible are they? In other words, what is the scope of this programme? Questions about processes must also be raised. Are educational processes the only ones available to the literacy worker, or can he also make use of organizational and distributive processes? Finally, what will be the outputs? Where will the impacts of the programme appear? Will those impacts appear in individuals, in groups, within institutions, or in the community as a whole?

DECISION-MAKING, CONTROL AND RESPONSIBILITY

The decision-makers involved in the change programme or functional literacy project must also be identified. How are they linked? How are the responsibilities of decision-making divided among them? What is the extent of the authority of each decision-maker at the various levels of the hierarchy? In other words, which decisions can be made by an individual and which are contingent upon decisions made elsewhere in the hierarchy? The answers to such questions will give evaluators an understanding of what kind of information will be needed, where and

by whom. Also, they will enable evaluators to develop a corresponding network for the team engaged in the evaluation of the functional literacy programme.

INTERFACES WITH OTHER CHANGE PROGRAMMES IN THE FIELD

A functional literacy project is frequently one component of the total social change effort launched in a particular community. The functional literacy project may be conducted by the Education Department, or the Community Development Department, but it may be based upon an agricultural theme. It may consequently assume that a parallel agricultural project is being implemented by the Department of Agriculture. Again, a functional literacy project may make assumptions about the existence of a nutrition education programme being conducted in the same area, and so on.

In such situations, the process of programme system articulation must ascertain how all the programme objectives are being shared; what interfaces between different programme systems have been built; and what means of integrating programme objectives have been proposed. Failing this, not only may the programme be impossible to evaluate, it may be difficult to implement in the first place.

Evaluation system definition

We have seen that, by extending the system metaphor to evaluation, it is possible to think in terms of input, process, output and context evaluation. It is also possible to develop a generalized model of evaluation. Since we cannot try to evaluate each and every conceivable aspect of a programme, we must develop situation-specific agendas, identify the network of evaluators and learn the golden rules of evaluation.

A GENERALIZED MODEL OF EVALUATION

As Figure 2 indicated, it is possible to apply the four system parameters to the evaluation system in order to develop a generalized model of evaluation. Theoretically speaking, one can evaluate an unlimited number of aspects of a programme or of relationships between variables entering the dynamics of change. One can ask questions about teachers, supervisors, administrators and community leaders. One can ask questions about funding procedures, administrative strategies, learning methods and communication techniques. One can

ask questions about impact—both long-term and short-term—with regard to changes in psychological attributes, attitudinal orientations, values, agricultural productivity and individual acceptance of new ideas.

No one programme needs to ask all these questions and most functional literacy projects have neither the time nor the money available to do so. When financial resources are limited, one should, indeed, consider whether or not the money earmarked for evaluation would be better spent on providing much-needed services. Evaluation programmes must grow out of well-considered evaluation policies and, in formulating evaluation policy, one must ask: What are the high-risk areas to be assessed for appropriate social action? What alternative strategies should be evaluated to economize on time and money? What questions must be answered to create a more sophisticated communication or distribution system? By eliminating a number of possible evaluation questions and retaining others, we eventually arrive at an evaluation programme.

A SITUATION-SPECIFIC EVALUATION SYSTEM

We have said that evaluation must help programme sponsors, planners and implementers to make choices. In other words, evaluation must service decision-making at all stages of a programme. A description of parameters and their variables will indicate what the decisions will relate to. Our knowledge of change theory and experience with similar work elsewhere will indicate the options for each decision, and our information needs for making those decisions, so that we can choose between available options. We must then design an evaluation system that will create the information enabling us to choose between options available in particular decision settings related to the inputs, processes, outputs and context of a programme of action.

Translating evaluation policy into a specific programme will entail: 1. Listing intended outputs of the programme and watching for any unintended or unanticipated consequences; 2. Choosing the points of impact where the change sought by the functional literacy programme should ultimately appear; 3. Considering the variations of contexts, inputs and processes and their combinations that might lead to more effective and efficient achievement of objectives; 4. Developing lists of indicators for use in determining the presence and intensity of impact; 5. Developing a generalized approach to evaluation that could accommodate a variety of evaluation models and techniques;

6. Deciding upon the nature of data to be collected in time series and the time intervals of this series; 7. Identifying the entire network of evaluators and clarifying the division of labour among all evaluators in that network; 8. Determining the points where data will be collected, collated, stored and analysed.

Let us now consider each of these points in somewhat greater detail:

1. What are the intended outputs of the programme being evaluated and what unintended or unanticipated consequences might it have?

The intended outputs of a typical functional literacy programme have already been listed. A particular project or programme, as we have previously indicated, may not be able to measure *all* possible outputs. Evaluation may, therefore, be limited to two, three or four important outputs: for example, literacy levels, productivity in farming, changes in group perceptions and establishment of modern economic institutions within the community.

In addition to their intended outputs, social change programmes, including functional literacy, may have unanticipated or unintended consequences, unfortunate by-products of the changes process. For example, emphasis on cash crops may depress food production and lead to malnutrition in families growing these crops. In order to recognize such consequences when they appear, evaluators must develop inclusive models and adopt open-ended approaches to evaluation.

2. What are the points of impact where change should become apparent?

Social change or functional literacy must make an impact on either physical units (agricultural tools, clothing, living space, fields) or social units (individuals, families, communities). A functional literacy team must decide, for the purpose of establishing an evaluation system, what units will be studied as targets of *impact*. One hypothetical project team might study only cotton-growing male farmers between 20 and 45 years of age and no other adult group in the communities. Another team might include wives of the adult farmers who are the primary target group. Evaluators may decide to study fields but not homes, credit co-operatives but not churches. The units to be studied as carriers of programme impact must, therefore, be selected.

3. What inputs, processes and contexts will best achieve the objectives of the functional literacy programme?

One major concern in building an evaluation system is the selection of crucial processes, interventions and their settings for study. Variations in inputs, processes and contexts must be tested for results. These variables must then be suitably differentiated in practice. In addition, the varied interventions must be standardized to validate conclusions about what intervention resulted in what impact, in what setting, and through what process.

4. What indicators could be used to determine the presence and intensity of impact?

The types of change we are looking for—new attitudes, new orientations, new social climates, new organizational patterns—are not visible to the naked eye. Physical impact is often directly observable, but socio-psychological impact (on individuals, groups and institutions) must usually be indirectly inferred. We cannot see intelligence, but a high or low score on an intelligence test might indicate the quality of a person's intelligence. In building an evaluation system, therefore, one must select suitable indicators of the desired change. It is possible that more than one indicator may be needed to study a single condition (of impact of a social unit). If social scientists disagree as to the best indicator of an inner condition, a set of complementary indicators may have to be used.

5. What evaluation designs will be used to generate needed information?

As has just been pointed out, socio-psychological phenomena cannot always be studied by direct observation and evaluation relies on indicators to demonstrate the existence and intensity of an inner condition of a social unit. Evaluation frequently involves comparisons in the status of the carriers of social impact before and after change interventions. Comparisons must also be made between groups that received or did not receive some intervention or sets of interventions, or between groups that received similar interventions in different settings undergoing different processes. In order to generate all this information, evaluation *designs* will have to be used. Part of the design problem is technical, but part stems from planning the evaluation system itself, and must be considered in the early stages of developing an evaluation system. Also, evaluation design must be capable of accommodating the study of unintended and unanticipated programme consequences.

6. What data will be collected in time series and at what intervals of time should such data be collected?

Data on adult learning and increases in productivity are obvious candidates for collection in time series, but there are some additional possibilities.

First, there is the question of base-line surveys, which are necessary if statements regarding impact and change are to be made in a 'before and after' format. Secondly, there is the question of phasing the programme, of determining intermediate objectives for each phase and of collecting evaluation data at the end of each phase. Thirdly, there is the question of long-term impacts of the programme after two, five, ten or twenty years. All these issues must be raised in the course of planning an evaluation system.

7. Who will the evaluators be and how will they perform their tasks?

Successful evaluation depends on a clear distribution of labour between external and internal evaluators, on the one hand, and between internal evaluators and programme staff, on the other. In building an evaluation system, decisions must be made about who will collect what data, where this data will be collated and stored, and who will analyse it for what purpose.

THE HIERARCHY OF EVALUATION NEEDS AND NETWORK OF EVALUATORS

Evaluation is not done by one man for another. Different people at different levels of decision-making need different kinds of information and, therefore, have different evaluation needs. There thus exists a hierarchy of evaluation needs. Identification of the various levels of decision-making and of the categories of information needed at each level of decision-making is necessary.

Again, while some people involved in a project have evaluation-related tasks, they should not be alone in performing evaluation. They should, of course, provide the initiatives for evaluation studies; design or assist in designing tests and observation schedules; train observers and field investigators; and finally, store and analyse data. However, they should never be expected to do this all by themselves. Every programme specialist must at some time play the role of evaluator. This may involve co-operation with the evaluators or form an independent component of the total evaluation effort.

THE GOLDEN RULES OF EVALUATION

Here are the two golden rules of evaluation: use the data you collect; and inform your informant.

Use the data you collect. Data collected at a particular decision-making level must first be used at that level of decision-making before being released to those at other levels. Evaluation is useless if information generated at a particular point does not improve implementation and action at that point. A literacy teacher who collects classroom attendance data to send to a supervisor must first use them himself. He must analyse them in relation to the life of his adult learners and ways of improving their learning. Similarly, before an area supervisor sends any information to district headquarters, he should use it to take deliberate actions with regard to his local programme.

Inform your informant. Those who inform must be told what has been done with the information they have provided. Thus, the area supervisor should tell his teacher how the other teachers are performing. District headquarters must tell each area supervisor what is happening in other areas, and so on.

9. What will be the points of data collection, data collation, data storage and data analysis?

The answer to this question will determine what type of data system will ultimately emerge.

Data system design

We can now move on to our next concern—building a data system to serve evaluation needs. The designing of the evaluation system and the data system are considered separately only for analytical purposes, since the two design tasks will, in fact, be undertaken simultaneously. The development of the data storage system will, indeed, contribute to the design of the evaluation system itself.

A programme will generate certain data as it is implemented; for example, information on annual budgets, number of teachers employed, number of adults enrolled, attendance figures, etc. The problem here is merely to *record* all this data in a suitably-designed register or tabulation. Naturally, not all the data generated by a programme will be numerical. Some will take the form of documentation, such as minutes of meetings, correspondence files, journals kept by teachers

and supervisors, log books, specimens of primers, sets of instructional materials, class schedules, supervisors' itineraries and reports, pictures, broadcast scripts, and suchlike. This type of data must also be carefully classified and stored before it is lost, as it is obviously an important component of any data system.

The data generated during the implementation of a programme is unlikely to be sufficient for all evaluation purposes. Additional data will have to be specially created by means of an *evaluation design*, which may involve naturalistic, quasi-experimental or experimental methods. Psychological tests and instruments may also be designed and administered. We shall discuss these topics in detail in later chapters.

To conclude, a data system for the evaluation of functional literacy programmes and projects must have at least two components, each of which may produce three types of data, as shown in Table 1.

Table 1. *Data for evaluation*

Type of data	Data generated during implementation	Data specially created
Numerical data	Class attendance	Changes in perceptions of adult learners
Qualitative data	Teacher's classroom journal	In-depth interviews with drop-outs from literacy groups
Documentation	Primers	A special newspaper story commissioned by the project

The following criteria may be useful in building a data system:

1. All needed information should be collected with the greatest economy of effort. Useless information should not be collected. Duplication in information collection is permissible only when an experiment is being repeated or the reliability of data is being checked.

2. While considerable amounts of information will have to be specially created at short notice by evaluators working on a functional literacy programme, a good data system should *anticipate* most programme information needs.

3. Data should be stored for easy retrieval and use. The evaluator should be able to aggregate data in various ways to suit his particular evaluation purposes. For example, if scores on attitudes towards 'one's own education' and towards 'the education of one's children' are available separately, in disaggregated form, an evaluator will be able to treat the two scores separately when developing suitable information. If the two scores are already aggregated as general attitudes towards education, the evaluator may find it impossible to disaggregate them into two separate scores.

4. The data system should be flexible. It should be possible to extend it with comparative ease or to adapt it to new data storage needs as they emerge.

5. One dimension that must be examined before constructing a data system is the *temporal* dimension. At what points in time will measurements be made to generate data? When programme activities fall into natural cycles and phases there is no problem, but in other cases time schedules will have to be drawn up specifically.

The evaluation system should have determined which variables of the programme system are to be studied. It should also have established what will constitute evidence of a process taking place, and what will be the indicators of impact on individuals, groups and communities. As indicated earlier, developing team agreements on the choice of indicators and on the nature of admissible evidence are an important part of developing an evaluation system. The design of the data system will complete the process of making programme or project evaluation operational.

ENTITIES THAT CHANGE

In our opinion, building a data system is mostly, *though not completely*, a matter of generating and storing for retrieval data about four entities or configurations[1]—individuals, groups, institutions and communities (sub-cultures). Additionally, we will need data on *artefacts* and *tools*

1. These are the four basic social configurations, and all change events can be analysed in terms of configurational relationships between one or more of them. In a broad impact programme of the functional literacy type, all four configurations will often be involved. See H. S. Bhola, 'Configurational theory of innovation diffusion', *Indian educational review*, vol. 2, no. 1 (January 1967). Again, H. S. Bhola, 'The methods and the materials of functional literacy', *Literacy discussion*, vol. 1, no. 4 (Autumn 1970) includes a brief discussion of the configurational model of change.

used in the programme. However, we suggest that, whether we are evaluating context, process, input or output, we will be concerned primarily with data on the four social configurations. Inputs in a functional literacy programme are individuals, groups and institutions; in some cases whole communities and sub-cultures. Thus, input evaluation must collect data on these configurations as well as on artefacts, tools and material inputs. Outputs in a functional literacy programme are these same configurations which have now been changed in a desirable direction because of the programme. While studying impact in output evaluation, therefore, we are concerned with these configurations in addition to physical/material entities such as seeds, agricultural tools, homes and fields. Processes can only be studied in terms of what is processed and how; that is, by observing what those processes do to those to whom they are applied. The processes of instruction, formation, extension, distribution, management and organization can thus be studied only in terms of what they do to the social configurations involved in a broad-impact programme, although some processes, of course, may be applied to physical entities, such as fields and homes. In context evaluation we are interested in the social, economic and political pressures on the environment, which are determined by sub-cultures undergoing a broad-impact programme and by the institutions located in those sub-cultures. Here again, although the physical and material components of the environment will have to be taken into consideration also, we are dealing with the four basic social configurations.

In conclusion, whatever the type of evaluation planned, the evaluator will require data on individuals, groups, institutions and communities (or sub-cultures). Additionally, he will need information on tools—and techniques and artefacts—used in development, but data on the social configurations are basic.

The following social configurations might be taken as sources for data collection, depending upon programme objectives:

1. *Individuals:* farmers; teachers; social leaders; supervisors; extension workers; administrators.
2. *Groups:* literacy classes; radio forums and discussion groups; demonstration groups; traditional groups.
3. *Institutions:* schools; co-operatives; local courts; party cells; families; factories; offices/departments/ministries.
4. *Sub-cultures:* communities; factions; castes.

To sum up, a data system will include numerical data, qualitative data and documentation. Some of this will become available during programme implementation if records of any kind are kept; some will have to be specially created by the use of evaluation designs. Most data will relate to the individuals, groups, institutions and communities (sub-cultures) involved, and the remainder, of course, to the tools, techniques and artefacts that those configurations employ to produce development.

A data system will have to design cards, tabulations and forms for recording data on selected configurations—individual farmers, groups, families, co-operatives, communities—which can be used to store data on various attributes of these configurations at various intervals of time, and to establish filing systems and a morgue for documentation.

It has been asserted that most or much of the data needed for *context* evaluation and *process* evaluation will have to be generated through specially designed instruments, using evaluation designs, although some may be generated as part of programme implementation. On the other hand, most or much of the data needed for *input* and *output* evaluation will become available through the implementation of the programme itself, although some input and output data may, again, have to be specially generated through evaluation design.

DESIGNING META-EVALUATION AND DATA SYSTEMS

In our foregoing discussion, we have indicated that a data system must reflect the needs of an evaluation system, which, in turn, must respond to the decision-making needs of a programme system.

Systems are, of course, composed of sub-systems and are themselves part of meta-systems. To take the example of the Unesco/UNDP Experimental World Literacy Programme, each country project could be considered as one programme system. However, a single country project might be working with different client groups, and using different processes of instruction and training to produce particular types of outputs in different regional contexts, so that country projects as systems may be composed of sub-systems. Again, the total Unesco programme, comprising twelve or more country projects, could be considered a meta-system with individual projects in different country settings as its sub-systems.

A meta-evaluation system that would service the decision-making needs of a total programme system, such as the Unesco Experimental World Literacy Programme, would have to be designed in the way we

have just described. It could not be left to emerge of itself from its various parts. It would, indeed, offer a complex exercise in programme definition and the clarification of decision-making needs, including needs for specific data at some central point.

In designing a meta-evaluation system, we would have to begin with a generalized model of the total programme system. Since a meta-evaluation system would deal with many country projects, the programme model would have to be generalized at an even higher level of abstraction. What kinds of decisions would be made by the decision-makers at the meta-system level? Once that question has been answered, evaluation policy, evaluation system design and data system design could be planned. The special problems faced in creating such a meta-evaluation system can be indicated here briefly. They arise from the need to develop a strong central 'nervous system' to communicate needs, receive data and monitor breakdowns; the need to develop a uniform vocabulary and to eradicate problems of definition; and, lastly, the need to develop instrumentation and procedures for interpretation so that generalizations can be made across projects and across cultures. All these tasks are by no means easy, but they can certainly be accomplished. Only questions that are never raised are impossible to answer.

From this general picture of evaluation, we shall, in the next chapter, turn to the specifics, beginning with the problems of measuring change in individuals, groups, institutions and communities.

Summary

This chapter has dealt with the problem of making evaluation *operational* in functional literacy programmes. The three major tasks involved were shown to be: (a) programme system articulation; (b) evaluation system definition; and (c) data system design. Just as the Chinese box in a box in a box cannot be opened without first opening the box outside the box outside the box, a data system cannot be built, nor data collected and stored, without first building an evaluation system; and an evaluation system cannot be conceived until the programme system and its decision needs have been analysed and understood.

The process of programme system articulation involves developing a generalized systems model for the programme and then articulating the means and ends relationships in a specific programme within a

particular setting at a particular time. The decision-making networks and interfaces of the functional literacy programme with other activities in the same community or region must also be identified.

Evaluation system definition also entails developing a generalized systems model and then defining a situation-specific evaluation model with a particular evaluation agenda. It is necessary to understand the hierarchy of evaluation needs of decision-makers at the various levels of decision-making and to identify evaluators, since evaluation is not the monopoly of those with evaluation-related designations. Two golden rules of evaluation were proposed: use the data you collect and inform your informant.

The data system, it was pointed out, would contain data that are spontaneously generated in the process of programme implementation and data that have to be specially created through evaluation design. Both types of data could be numerical or qualitative or take the form of documentation. The entities on which data would be collected were identified as individuals, groups, institutions and communities (sub-cultures). Some problems of developing meta-evaluation and data systems were discussed.

CHAPTER SIX

Measurement of change: the classical protocols

Evaluation sets values on attributes and achievements and thus involves measurements. These may be quantitative or qualitative but must always be objective.

Today, methods and procedures of quantitative measurement form part of the tool kit of most researchers and evaluators. The most usual approach involves setting up learning or change objectives, selecting indicators and scales, developing and validating test items, checking the reliability of tests, and collecting and analysing measurement data. In this monograph we shall refer to these procedures as the *classical protocols* because they follow strict established rules and are accepted as authoritative.

While these classical protocols have, to some extent, standardized the process of measurement in education, they have also created certain problems of their own. Some educators believe that these approaches allow them to measure definitively: sadly, they have often measured trivialities. In trying to make problems amenable to scientific testing, they have artificially segmented them into minute parts, knowledge of which normally adds little to our understanding of reality.

The critics of the classical protocols have recommended the use of qualitative and holistic methods in the measurement of change. They suggest that educators use the measurement methods of the historian, the anthropologist and the naturalist. They propose participative approaches. These approaches are by no means new, but have certainly been rediscovered during the last decade. These rediscovered protocols will form the subject of Chapter Seven.

For the time being, we shall concentrate on the classical approach to the measurement of change. The concept of indicators will be introduced, criteria and norms for measurement and testing will be discussed and scales and instruments for measuring individual and

social attributes of particular interest to functional literacy workers will be examined in detail.

The perennial evaluation question is: Did things change? Since systematic evaluation seeks to present impressions objectively, we often ask the related questions: If things changed, how can one demonstrate that they did? In which direction did they change, and by how much? As objectivity depends on accuracy, our answers must be based on exact measurements.

While discussing ways of making evaluation operational, we stated that social change can be seen most frequently in individuals, groups, institutions and communities (and occasionally in the tools and artefacts that men use). Now, when we wish to measure social change, we discover that the psychological states of individuals, sociological conditions of groups, organizational aspects of institutions and cultural patterns of communities are not visible to the naked eye. We can count the number of adults in a literacy class but we cannot see their motivations for being there. We can count the number of people who drop out of classes, but cannot discern for what reasons. We can observe a group of farmers listening attentively to the agricultural extension worker but not how they perceive the speaker, nor how group attitudes may be changing. We can count an organization's employees and study its organizational chart or organigram, but we cannot observe the quality of its decisions or its effectiveness in providing services. Similarly, one cannot see community ethos, structures, cohesiveness or conflict; all are invisible.

Indicators of human and social conditions

The human and social conditions mentioned above, invisible to the naked eye, can be observed only through their indicators. We assume that the regularity with which adults come to literacy classes is an indicator of their motivation to learn. We assume that scores in a reading test are indicators of reading ability and that higher reading ability is indicated by higher scores. Similarly, we consider a particular pattern of answers in an interview to be an indicator of change in individual values and attitudes. We believe that the cohesiveness of a group may be indicated by the number of positive interactions between group members at different times. We accept a profile of responses by employees of an organization as an indicator of that organization's efficiency. We consider that the establishment of a

co-operative and a bank, expenditures on agricultural inputs and patterns of consumption by families may indicate that a community is on the road to modernization.

INDICATORS AND THEORIES

Indicators arise from theories. Sometimes our theories are simple and stand us in good stead; for example, a series of appropriate questions by learners in a group is a fairly reliable indicator of group interest. At other times we may be concerned, for instance, with change in individual values or in cultural norms. We may theorize about the nature of values, how they change and how changes can be registered by means of tests, interviews or observation of behaviour. We may theorize, again, about how different profiles of responses might indicate different degrees of change in values, norms and attitudes. Thus our theories may develop into an extremely complex network.

Indicators, like the theories from which they are derived, may be good, bad or spurious. We may think that a higher level of questioning activity in a group is an indicator of group interest, while it may be, in fact, an indicator of fear and anxiety. We may consider regular class attendance to be an indicator of learners' motivation to learn to read, but, in reality, it may be an indicator of their desire for social contact in a group setting not otherwise accessible to them. We may think that participation by different community factions in a planning group is an indicator of a new spirit of co-operation, but it may result only from a desire to protect factional interests and perpetuate the *status quo*.

Indicators can be interpreted at different levels. Fewer visits to the health clinic may be an indicator of improved community health, which may in turn be an indicator of the success of the health extension and nutrition education programmes in the community. On the other hand, a decrease in the number of visits to the health clinic could also indicate that the clinic is inefficient or that the traditional medicine man has reasserted his influence. It is, therefore, important for literacy workers to use more than one indicator of a psychological or socio-cultural condition wherever possible, in order to ensure the validity of their conclusions.

SOME INDICATORS OF THE EFFECTIVENESS OF LITERACY PROGRAMMES

In the preceding section, we provided examples of indicators that a literacy worker engaged in evaluation might encounter. He would, of course, have other indicators to consider when developing a literacy

evaluation system. The Unesco Experimental World Literacy Programme worked with a set of indicators that could usefully be examined by all those engaged in literacy evaluation. The Panel for the Evaluation of Experimental Literacy Projects, set up by Unesco in one of their meetings in Tehran, Iran, during September 1970, drew up a list, reproduced on the following pages, of eleven minimum and nineteen recommended indicators for use by the various projects.

LIST OF MINIMUM AND RECOMMENDED INDICATORS

I BASIC STATISTICAL INDICATORS

A. Quantitative aspects
   1. *Minimum indicators*
      1.1. Ratio of inscription in programme: Indicator is a ratio between the number of participants originally registered and number of places available.
      1.2. Rate of drop-outs
      1.3. Rate of attendance
      1.4. Time utilization: Indicator is a ratio between class sessions planned and actually held.
   2. *Recommended indicators*
      2.1. Rate of coverage: Indicator is a ratio between number of eligible clients and those actually covered in a programme.
      2.2. Rate of participation in final tests

B. Qualitative aspects
   1. *Minimum indicators*
      1.5. Literacy acquisition
      1.6. Acquisition of technical and professional knowledge
   2. *Recommended indicator*
      2.3. Acquisition of knowledge of socio-economic character

C. Degree of adoption of writing, reading and calculating
   1. *Minimum indicator*
      1.7. Use of writing ability
   2. *Recommended indicators*
      2.4. Use of reading ability
      2.5. Use of arithmetical ability

II INDICATORS PERTAINING TO ECONOMIC GROWTH AND DEVELOPMENT

A. Production
   1. *Minimum indicator*
      1.8. Increase in production per capita

2. *Recommended indicators*
   2.6. Quality of products
   2.7. Selling price
   2.8. Price (per quantitative unit) of the elements entering the costs of production

B. Income—living standards
   1. *Minimum indicator*
      1.9. Increase in the number of durable goods and improvements contributing to the standard of living
   2. *Recommended indicator*
      2.9. Increase of net global monetary income of individuals

C. Production, servicing and transport equipment
   2. *Recommended indicator*
      2.10. Increase in equipment for production, servicing or transport (in programmes for small agricultural producers)

D. Socio-economic attitudes
   2. *Recommended indicator*
      2.11. Importance of changes in the domain of socio-economic attitudes and in the individual's role in society.

III ATTITUDES TOWARDS EDUCATION

1. *Minimum indicator*
   1.10. Rate of scholarization of participants' children: This is a ratio between the total number of children of school age and those actually attending school.
2. *Recommended indicator*
   2.12. Rate of interruption of schooling in participants' children: Ratio between the number of participants' children who interrupted their schooling and of interruptions in the total school-going population.

IV PROFESSIONAL COMPETENCE

2. *Recommended indicators*
   2.13. Adoption rate of recommended innovations
   2.14. Level of know-how in the fields covered by the programme
   2.15. Increase in desire for changes and technical innovations

V RELATION TOWARDS MEANS OF MASS COMMUNICATION

2. *Recommended indicators*
   2.16. Ownership of radios and television sets
   2.17. Preference for educational programmes

VI HEALTH, HYGIENE AND SAFETY

1. *Minimum indicator*
   1.11. Acquisition of knowledge of health, hygiene and safety around home
2. *Recommended indicator*
   2.18. Acquisition of knowledge of health, hygiene and safety in work settings

VII COSTS AND COST-BENEFIT ANALYSIS

2. *Recommended indicator*
   2.19. Per capita cost of functional literacy

Some features of this list of indicators should be noted. First, it emerged from Unesco's concept of functional literacy and the objectives of functional literacy programmes in both rural and urban settings. A more generalized model of literacy in development would have led to a somewhat different set of indicators. Secondly, the emphasis seems to be on indicators of inputs and outputs. Outputs are viewed in terms of information and attitudes learned and an attempt is made to measure them in terms of the *applications* of the information learned and the *impacts* of such applications on the life and work of participants. Thirdly, indicators have been divided into two clusters—minimum and recommended. The implication is that, while it might be advisable for a functional literacy project to use all of the thirty indicators suggested by the Unesco Panel, conditions will differ from programme to programme and from country to country, so that each project will have to develop its own evaluation policy. Fourthly, most of the indicators listed are not themselves directly observable and will have to be defined in terms of others that can be measured and weighed.

Developing a set of indicators for use in the evaluation of a functional literacy programme is a complex task. It should be undertaken by the entire project team in close co-operation. The indicators selected, in addition to being theoretically sound, should be sufficiently objective to ensure that different people observing the same change at different times would describe and measure it in identical terms. To guarantee the validity of an indicator, other indicators of the same condition should be developed and used to provide support.

Criteria and norms for measurement

Once we have chosen the observable indicators of unobservable individual and social conditions, two further problems arise: that of defining norms and criteria for judgements and comparisons and that of designing measuring instruments.

We may consider scores on a reading test to be an indicator of a learner's reading ability. But what norms and criteria can we use to separate the illiterate from the semi-literate, and the semi-literate from the fully literate? How can we separate good readers from readers who are not very good? Should verbal responses on an attitude test be used alone or in conjunction with behavioural criteria? What criteria should be used to separate the early adopter from the late adopter of agricultural innovations? What level of performance with new tools indicates satisfactory learning of technical skills? What are the criteria for proper co-operation within groups and healthy climates within organizations?

Questions of norms and criteria frequently arise during the process of developing indicators. Criteria must be tested in real-life situations and modified if need be. Experience may show that criteria based on verbal scores fail to anticipate actual behaviour. Functional literacy criteria may have to be altered to include a particular level of computational ability, and so on. Criteria for selecting literacy teachers for training or for judging the effectiveness of teacher training may prove unsatisfactory. Criteria should, therefore, be developed thoughtfully and systematically, and must themselves be tested and revised whenever necessary.

CRITERION-REFERENCED AND NORM-REFERENCED TESTS

For years, schools have been testing students and then assigning them marks. They have compared one student with another, one class section with another, and one particular grade with that preceding or following it. These or similar questions have been asked: Does this student read well? Does this student read better than other students in the same class? Do students in one class read as well as students in another class at the same grade in the same school? How does the reading ability of fourth-grade students in one school compare with that of fourth-graders in schools in other parts of the state or in other states? In order to make such comparisons, national reading ability norms must be known. One can then compare a student with the norm

and say that he reads at the fourth- or fifth- or eighth-grade level.

Norms are established essentially as follows. A test is designed and is administered to a large sample of, say, all third-graders in a state, region or country. The answers are marked and the test is successively revised on the basis of feedback to ensure that most third-grade students in most of the nation's schools are able to answer most of the test items correctly. The test then becomes a standardized test for that grade and is called 'norm-referenced'.

Test-makers do not always consider it necessary to develop norm-referenced tests. They want to judge students in terms of learning objectives according to specific criteria for success in the attainment of those objectives. For example, functional literacy workers could agree upon a set of behaviours or attitudes to be learned by farmers in a community and decide what criteria would be used to judge mastery of learning. If a farmer can learn all those attitudes and behaviours and attain the pre-determined level of performance, he is declared a success. No comparisons are made between this farmer and a 'typical' farmer. Such tests are called 'criterion-referenced' tests.

Functional literacy workers normally—and properly—use criterion-referenced tests of reading, attitudinal changes and economic skills. However, they are often requested to compare the level of literacy their programme is producing with reading ability as graded in the nation's schools. They are persistently asked: Do your adult learners read at the fourth-grade level? Is a particular grade level sufficient to avoid relapse into illiteracy? Literacy workers may be indifferent or hostile to norm-referenced tests, but they will frequently be questioned in normative terms.

The best way of coping with this situation is to continue testing the achievement of programme objectives by means of criterion-referenced tests, but to obtain the fourth- to eighth-grade textbooks used in the school system for use as unseen reading tests to obtain some idea of equivalent reading levels in literacy classes and schools.

Scales and instruments of measurement of interest to functional literacy workers[1]

We have indicated above that functional literacy workers do not deal with conventional measurements. They have to design their own *scales*

1. This material was originally collected by Dr. John J. Hanitchak and later revised by David P. Rutt, Jr., and the present author.

(in place of lengths, weights and volumes) and develop their own *instruments* (in place of yardsticks, balances and calibrated cylinders).

In this section we shall describe the four types of scales—nominal, ordinal, interval and ratio—most often used in literacy evaluation. We shall then discuss the instruments used in conjunction with these scales: questionnaires, interviews, observation sheets and various tests. We shall outline the design and use of these instruments in general terms, enumerating their special strengths and weaknesses, and list the precautions to be taken when using them.

TAXONOMIES OF OBJECTIVES

Before presenting the various scales of measurement, let us first consider taxonomies of instructional (or change) objectives. It can happen that, in testing for information learned, attitudes or values internalized or skills acquired, we become confused about objectives. Having taught with one objective in mind, we may proceed to test for something not included in the teaching objectives.

Professor Benjamin S. Bloom and his associates have suggested that within the *cognitive domain*, instructional objectives can be classified as shown in the following pages. Similarly, instructional objectives in the *affective domain* (attitude change and values) could vary as shown in the outline. A taxonomy of instructional objectives in the *psychomotor domain* has recently become available, but its usefulness for literacy workers is limited.

Even a cursory look at these taxonomies should give rise to the objection that information does not always ensure real comprehension and comprehension does not necessarily lead to application, analysis or evaluation. Similarly, it is possible to respond positively to particular values and attitudes without genuine commitment and to have a set of discrete values that do not add up to an organized value system. We must be very clear about what we are teaching and what we are testing for.

INSTRUCTIONAL OBJECTIVES IN THE COGNITIVE DOMAIN

1.00 Knowledge
- 1.10 *Knowledge of specifics*
  - 1.11 Knowledge of terminology
  - 1.12 Knowledge of specific facts
- 1.20 *Knowledge of ways and means of dealing with specifics*

- 1.21 Knowledge of conventions
- 1.22 Knowledge of sequences
- 1.23 Knowledge of classifications and categories
- 1.24 Knowledge of criteria
- 1.25 Knowledge of methodology
- 1.30 *Knowledge of the universals and abstractions in a field*
- 1.31 Knowledge of principles and generalizations
- 1.32 Knowledge of theories and structures

2.00 Comprehension
- 2.10 *Translation*
- 2.20 *Interpretation*
- 2.30 *Extrapolation*

3.00 Application

4.00 Analysis
- 4.10 *Analysis of elements*
- 4.20 *Analysis of relationships*
- 4.30 *Analysis of organizational principles*

5.00 Synthesis
- 5.10 *Production of a unique communication*
- 5.20 *Production of a plan, or proposed set of operations*
- 5.30 *Derivation of a set of abstract relations*

6.00 Evaluation
- 6.10 *Judgements in terms of internal evidence*
- 6.20 *Judgements in terms of external criteria*

INSTRUCTIONAL OBJECTIVES IN THE AFFECTIVE DOMAIN

1.00 Receiving (attending)
- 1.1 *Awareness*
- 1.2 *Willingness to receive*
- 1.3 *Controlled or selected attention*

2.00 Responding
- 2.1 *Acquiescence in responding*
- 2.2 *Willingness to respond*
- 2.3 *Satisfaction in response*

3.00 Valuing
- 3.1 *Acceptance of a value*
- 3.2 *Preference for a value*
- 3.3 *Commitment*

4.00 Organization
- 4.1 *Conceptualizing a value*
- 4.2 *Organizing a value system*

5.00 Characterization by a value or value complex
- 5.1 *Generalized set*
- 5.2 *Characterization*

SCALES OF MEASUREMENT

We have already talked about the use of indicators to identify changes in individuals, groups, institutions and communities. Change, as you will recall, is the difference shown by an indicator in the subject between two points in time. By measuring this difference we can determine how far a student is progressing in his attainment of various skills or whether the attitudes of a group are becoming more positive. It is generally not enough to say that something or someone is 'better' or 'worse' than before. We must be able to state *how much* of a difference has taken place and in what direction.

Measurement can be defined as the assignment of numbers or values to objects or events according to rules. An example may clarify this definition. You have been put in charge of a foot race and are required to determine the five best runners. You have chosen to assign places (numbers) to the runners (objects) according to the order in which they finish the race (rule). Thus, the runner who finishes the race before anyone else will be first, the next one will be second, and so on. Unfortunately, when we wish to measure the level of motivation or achievement of an adult attending literacy classes our task is more complex. We are still assigning a value or number to an object according to a rule, but in this case the object is a pattern of response that the adult makes on a test. This, as we shall see, can present problems.

Measurement is a part of our daily lives. We have a clock to measure time, a ruler to measure distance and scales to measure weight. Clocks, rulers and scales are all useful and enable us to apply numbers to objects in our everyday world. However, what happens when we try to deal with human characteristics or social phenomena? What is personality and how can we measure it? When is a person literate rather than semi-literate? These questions are important in a literacy programme, yet we have no ready-made rules or scales to help us answer them.

In attempting to measure behavioural aspects of functional literacy programmes—such as achievement, attitudes and application of skills—we are faced with two major problems. The first is that of devising 'good' indicators. The second is that of selecting measurement instruments (questionnaires, interviews, observation schedules and tests) that will enable us to 'apply' values to these and other concepts accurately and consistently. It should be obvious that both indicators and measurement instruments must be well designed if we are to gather any worthwhile information for our decision-making.

It may have occurred to you that measurements are not necessarily absolutes. We can say either that a man is 'short' or that he is 5 feet 8 inches (1·7 metres) tall. While both of these statements are true, they differ as to their precision because they use different standards of measurement. We shall now examine various scales of measurement.

The nominal scale

This scale classifies objects in such a way that all the objects within one category are similar with respect to some property. The classification of all adults in a community into males and females, or young and old, is an example of nominal measurement. Here we are assigning not numbers, but categories. Of course, we could classify people according to, say, occupations and give each category a number rather than a name, but numbers at this level would have *no* significance other than to distinguish between categories; if we assign the number '1' to farmers, and the number '0' to fishermen, this does not mean that farmers are 'greater' than fishermen. It is important to keep this in mind, because numbers used in the nominal scale indicate *nothing* about the value of the categories, except that they are different.

In this scale, the criteria used to define a category are important. Some categories are straightforward: for example, sex, age or occupation. Other categories—such as literate/illiterate, high achiever/low achiever—can present difficulties. Is a person literate when he can sign his name or when he can pronounce the words on a written page? We can agree and/or communicate about categories only so long as we can agree upon the criteria used to place individuals in those categories.

The ordinal scale

The ordinal scale employs nominal classification, but also orders the categories in relation to each other. For instance, community youth athletic teams can be categorized as first, second, third, etc., according to some rule or criterion, most probably the number of contests won. We can judge the work of one village craftsman as being better than that of another on some basis of artistic merit. We can classify adult learners with respect to their literacy achievement. Evaluators often categorize the attitudes of adult learners as 'good', 'medium' or 'poor'. Thus, with the ordinal scale we are saying that entities within one category have more or less of some characteristic than entities in

another category. We recognize that a difference exists although we do not say how much of a difference.

In attempting to categorize learners' attitudes towards the literacy class, a literacy teacher may decide that a useful and easily-utilized indicator of attitudes would be attendance at class sessions. He or she may go even further and state that perfect attendance or one absence would represent a 'good' attitude towards the literacy class, two or three absences a 'medium or average' attitude, and four or more absences a 'poor' attitude towards the class. If, knowing these criteria, we study the literacy teachers' records and find that adult 'A' has a 'poor' attitude, we readily know what this means.

This scale appears to be both objective and logical. Nevertheless, from the standpoint of validity we must ask whether class attendance really reflects an attitude towards the class. An adult may have very positive feelings towards the class and yet be prevented from attending regularly by personal difficulties.

The interval scale

The third standard of measurement, while possessing the characteristics of the two previous scales (ranking within a defined category), uses, in addition, numerical units of measurement that have *equal intervals*. This means that the difference between successive units in any part of the scale is always the same. For example, the distance between 6 and 8 centimetres is equal to the distance between 25 and 27 centimetres. Similarly, the difference between 25 and 26 degrees and between 14 and 15 degrees on a thermometer is the same in terms of the change in heat.

A literacy test, however, is neither a ruler nor a thermometer. What can we learn from two test scores? Let us say that learner 'A' scored 45 out of 50 in a test, while learner 'B' scored 35 out of 50. Normally test scores can be treated as ordinal data, so that we may state with confidence that 'A scored better than B'. Treating these scores as interval data we can also say that 'A scored 10 points better than B'. However, if learner 'C' had scored 20 on the test we could *not* say that 'A' had achieved more than twice as much as 'C', but only that he had scored 25 more points, as we shall see in the following paragraph.

The ratio scale

This standard of measurement has the three characteristics of the interval scale, plus a fourth: an *absolute zero*. This permits us to relate proportionately different measurements. For example, we are able to say that 2 metres is twice 1 metre and 4 kilos represent a quarter of 16 kilos only when we have an absolute zero. Now, although zero weight and zero length are conceivable, zero achievement is not. This explains why, in the preceding section, we could not say that learner 'A' had achieved more than twice as much as learner 'C'. The latter learned *something* during a given unit of instruction, however little. The fact that it does not show up may be a fault more of the test than of the learner.

The important point emerging from this discussion is that numbers can be used in different ways, and that the information obtained will vary accordingly. The measurement procedures available to us in the area of education are relatively crude. Employing the nominal scale, we use categories such as 'pass' or 'fail', and we rank within categories by grades or levels of attainment. We can occasionally use an interval scale if, for example, we wish to calculate averages on test grades but the information obtained must be interpreted and used with caution. It is therefore wise to maintain a level of scepticism with respect to the significance of numbers derived from the measurement of behaviour. These numbers should be used only in conjunction with other types of information obtained from observation, questionnaires and interviews.

We have stressed the difficulty of measuring learners' knowledge, change in attitudes or utilization of newly-learned skills, pointing out the necessity of selecting reliable *indicators* that will enable us to infer that certain changes have taken place.

Measuring the changes that have taken place calls for the development of a variety of measurement instruments: questionnaires, interviews, observation schedules and tests. In the following section we shall explain the use, development and construction of these instruments.

INSTRUMENTS OF MEASUREMENT

Questionnaires. The questionnaire is probably one of the most widely used—and misused—tools for the collection of all types of evaluation information. A questionnaire, as its name implies, is a written list of questions given to a person who is asked to respond to them. Responses may be either 'closed', that is, the respondent can select

only from given responses, or 'open-ended', allowing the respondent to answer however he wishes.

The type of information that can be obtained via a questionnaire is often factual. For example, the following questions might be asked: How many children do you have? How long have you lived in this town? How many, and what types, of jobs have you held in the last five years?

Before the questionnaire is chosen as a means of collecting data, its advantages and disadvantages must be considered. Questionnaires are relatively inexpensive to develop and use. They allow large quantities of information to be collected within a fairly limited time period with little involvement of personnel. The use of 'closed' responses increases reliability. The privacy connected with the administration of a questionnaire normally ensures that respondents answer truthfully and unreservedly.

Nevertheless, these benefits are counterbalanced by the following drawbacks. First, questions may be interpreted differently by individual respondents. Secondly, the time and care required to complete a questionnaire demand high motivation and interest on the part of respondents. Thirdly, a return of 50 per cent of the questionnaires sent out can be considered very good, but does not necessarily give a true reflection of reality; for instance, if *only* the educated persons in a community respond to your solicitations for information, you could have a seriously biased picture of community needs. Fourthly, the use of open-ended responses results in problems of analysis and interpretation. Finally—and this has implications for literacy programmes—questionnaires cannot be used with illiterate groups since reading and understanding are involved.

The construction of questionnaires requires caution. One must have a clear idea both of the information needs and of the audience *for* whom the information is being collected and *from* whom it is being gathered. It is not sufficient to ask about 'community needs' in some vague way. One has to determine specific categories such as 'educational needs', 'recreational needs', 'health needs', etc., and define, within these categories, the kinds of information that would be helpful.

The purpose for which information is needed is all-important. You may wish to know what forms of basic education are currently available in the community in order to assess possible gaps and find out what existing resources could be drawn upon to offer more comprehensive programmes. In this case, your questionnaire should be directed toward all

relevant sources of information. It makes little sense to solicit an individual or group that either knows nothing about the programme or has no interest in it.

Since one usually has to rely upon the goodwill, understanding and motivation of respondents, a questionnaire must be designed in such a way that it can be answered easily, with the number of questions and the time required to answer them kept within reasonable limits. This necessitates spending considerable time on the actual design of the questions themselves.

When writing questions, you should concern yourself, first, with their relevance to your purpose for collecting information. Questions should not be ambiguous nor should they lead the respondent into giving the 'right' answer. You should beware of asking for information a person simply may not possess. Finally, try to avoid questions that imply social desirability, such as 'Do you think education is good?' or 'Should we build more libraries?'

Wherever possible, the use of closed (check-off) items rather than open-ended items is advisable and will avoid many of the problems mentioned above. Such an approach will help to clarify the type of information you are seeking, decrease the time needed to respond and help to standardize the responses, thereby making easier the analysis and interpretation of the information received. For example, an 'open-ended' question might ask: 'What is the sex and age of your adult learners?' Alternatively, you could construct the question as follows:

Record the number of learners in your programme in the spaces below:

Male 15–19 years _____	*Female* 15–19 years _____
20–24 years _____	20–24 years _____
25–34 years _____	25–34 years _____
35–44 years _____	35–44 years _____
45 and over _____	45 and over _____

The latter question is easier to answer and is already coded into the response categories you will need.

Similarly, you could ask: 'What do you consider to be the priority educational needs of adults in this community?' Such an open-ended question would probably bring a veritable deluge of suggestions. It would be preferable to elicit the information required as follows:

> Below is a list of possible educational needs of the adults in your community. Indicate whether you believe the items have a *high* or a *low*

priority for meeting your community's needs. Place an *H* for high or an *L* for low opposite each.

A. Basic literacy _____ D. Use of leisure time _____
B. Health information _____ E. Occupational training _____
C. Family planning _____ F. Other (write in) _____

If you were in need of more specific information or a finer comparison you could ask your respondents to rank each of the above items, using 1 for highest priority, 2 for next highest, 3 for third highest, and so on. In either case, you have standardized the language and obtained material ready for analysis.

The order of items within a questionnaire is also important. If information of a personal nature is sought (income, religious views, etc.), it may be advisable to reserve such queries until the end of the questionnaire, starting with questions of a factual nature that can be answered easily and do not alarm or annoy the respondent. Remember, however, that caution must always be used when asking for information of a personal nature.

Administering a questionnaire is just as important as planning and writing it. Before attempting to answer a questionnaire, a respondent should be told who is collecting the information, how it is to be used, why it is needed and whether his responses will be made public or kept confidential. Clear instructions for completing the questionnaire should also be given, for example:

> As a part of the Ministry of Education, the Department of Basic Adult Education is interested in providing adult illiterates with opportunities to learn the basic skills of reading, writing and arithmetic. In addition, where possible, we would like to provide programmes that will benefit them in their work, health and family life. In order to plan realistic programmes we need information . . .
>
> Because the resources of our country are limited, we want to co-ordinate and co-operate with others interested in this task . .
>
> The information is being collected from . . . and will be used only by personnel of this office. Our interest is in adults who are engaged in formal studies. Our reference points are males and females who are fifteen years or older and not enrolled in any educational programme at this time.

Any such statements, of course, should use language appropriate to the levels of understanding of the respondents.

Interpreting the data obtained from the questionnaire must be given due consideration. With the first example given above (age and sex of learners), responses can be accepted at their face value. On the other

hand, if you were to ask 'How many adults in your community are in need of basic education?' respondents might pick a number 'out of the blue', give you data from an actual census, or misunderstand the meaning of 'community'. You would, therefore, need to know on what basis estimates were made before using the figures given.

Similarly, responses concerning the priorities of education needs in a community should be viewed with a healthy scepticism. A listing of priorities may be based upon subjective, biased opinions or an objective economic analysis of the community. Knowing something about the respondents from whom you receive information can be helpful in interpreting data or deciding how much confidence to place in their judgement. Accordingly, it is helpful, if not essential, to include in your questionnaire a request for information about the person answering the questions. Not all people willingly provide personal data and some may refuse to do so, so that much will depend upon how your request is worded and the clarity of the reasons given. It may be a good idea to test the questionnaire with a sample group and use feedback to make needed alterations before administering it to the intended respondents.

Interviews. The interview is basically a face-to-face version of the questionnaire. An 'interviewer' questions a respondent to obtain answers that are relevant to the purpose of a certain programme. Like the questionnaire, the interview can elicit needed facts: when, where, how much, etc. The one thing an interview can do that a questionnaire cannot is to ask the question 'why?', in other words, to explore a certain topic in depth to reveal the reasons and motivations that made a person act in a certain way. This is without doubt the major advantage of the interview. It gives you the opportunity to probe in depth and clarify areas where ambiguity and misunderstanding may exist.

The great disadvantage of this method is the amount of time involved. It can take many long hours to get information from an individual or to follow up a previous interview for more details. Thus the interview is not the most economical method of gathering information in terms of time or money.

The preparation of an interview is similar to that of a questionnaire. The information you desire should be known and organized ahead of time and an interview guide developed for use by the interviewer, to make sure he covers all topics of interest. The amount of detail will

depend upon your reasons for using this technique. If you have chosen the interview because your respondents cannot read, you are not able to deliver questionnaires, or you need a 100 per cent return if possible, your interview guide may well be as detailed as a questionnaire. On the other hand, if you are interested in broad, original, in-depth thinking on the part of your subjects, your guide may simply list the main topics you wish to cover.

The critical variable in interviewing is the interviewer himself. He is the mediator of the information and his presence and behaviour can serve to inhibit or facilitate response. His own interpretations and biases may creep into the information. Consequently, the selection and training of interviewers is obviously an important concern. One must select individuals who are not only stable, but also acceptable to respondents. Interviewers should have a thorough understanding of the project and its purposes, so that they can offer explanations and also realize the significance of the material obtained. Some consistency of method is advisable and can best be accomplished through group training, giving interviewers opportunities to observe and be observed, and to criticize each other. It is, however, impossible to standardize selection and training as so much depends upon the context of the interviews, the material and the accessibility of the respondents.

As with questionnaires, it is best to organize trial interviews so that any difficulties with interviewers or interview guides can be studied and eliminated.

Observation schedules. Evaluators sometimes have no questions to ask; they simply wish to observe the behaviour of individuals or changes in their fields and around their homes. The suggestions made above with regard to the design of questionnaires and interview guides can be applied directly to the design of observation schedules. The observer must go to the field fully prepared, knowing what he or she will observe and in what form the observation will be recorded. We will return to observation procedures in Chapter VII.

Tests. Tests are a useful way of creating and collecting information about the knowledge a learner has acquired within a given unit of instruction. Testing permits us to assign numbers to a learner's achievement according to definite rules. By assigning marks to learner responses we can make inferences about what has been learned. Nevertheless, despite the apparent simplicity of this method, many

complex questions are involved. Certain procedures must be strictly adhered to so that the inferences made reflect what the learner 'really' knows, as accurately as possible.

Tests can be constructed to measure reading and arithmetic skills, writing ability or knowledge of scientific concepts. Most of the tests you will use will probably be constructed by yourself or by a team of teachers. They will thus differ from 'standardized tests', which are usually administered to thousands of people so that standards or norms of performance can be established. By administering such a test to members of a class, it is possible to compare the performance of that class with that of the larger population and consequently to determine the level of achievement or predict the potential ability of students and groups. These types of standardized tests are called 'achievement' and 'aptitude' tests respectively, but they are of little use to the functional literacy worker.

As with any other data collection instrument, care must be taken in the design, construction and administration of a test. We will examine each of these three topics in detail on account of the importance of testing in literacy programmes.

TEST DESIGN

In *designing* a test, the following guidelines should be kept in mind:

1. The first step is to define the desired change clearly, either mentally or in writing. Each instructional unit has no doubt presented important ideas, concepts and principles for the learner to absorb. It is, therefore, helpful to set definite objectives for the learner's performance, and to select clear and unambiguous indicators enabling you to infer that these ideas, concepts and principles have been grasped. It is important to remember the relative emphasis that different objectives have received throughout the unit of instruction, so that the test will reflect that same emphasis.
2. Next, you must determine the standards of success that will be used in judging the learner's performance. It is possible to set either an absolute standard or a relative standard. In the first case, an absolute standard of success is set ahead of time (e.g., the learner will answer 8 out of 10 questions correctly). All learners scoring at or above this standard are then judged to have mastered at least the minimum level of learning for that unit. In the second case, the test is administered to a class and levels of success are established in

accordance with the distribution of scores within the class. You are then comparing the learners with one another to determine their success or failure in a particular unit of study.

At times it may be necessary to know how one learner compares with others in the adult group. Generally, however, we are interested less in relative standing than in the actual performance of a learner with reference to levels of achievement considered important by the programme. Knowing that a learner scores high or low in relation to some hypothetical group is less important than knowing whether he scores high or low in relation to what has been taught. Accordingly, we tend to use criterion-referenced, not standardized, tests to see whether performance satisfies the criteria for success.

3. Lastly, you must decide whether testing will make use of objective or subjective questions. The objective test includes multiple-choice true/false and matching items; the subjective type includes essay, short answer and completion questions.

   The basic difference between the two lies in the type of response required and the means used to assess the 'correctness' of the answer. Subjective questions allow for an open-ended response, enabling the learner to answer in his own words. Obviously, answers may also be subjectively interpreted by whoever grades them; the degree of 'correctness' of an answer is thus a function of the grader. Objective test items require the learner to select from a given set of responses, which have been compiled ahead of time, so that the test will always be graded in the same way.

   Prime considerations in deciding which types of test to use will be construction and grading. Objective-type questions take time to construct but are easy to grade. On the other hand, subjective-type questions are relatively easy to construct, but time-consuming to grade, with many opportunities for personal bias to affect the grading procedure.

We will now examine a number of specific test formats, proceeding from the most subjective to the most objective.

TEST FORMATS

Essay. This is the easiest type of test to write, although certain precautions are necessary. An item such as 'Discuss democracy' is much too broad, vague and open-ended. 'Discuss the role of a

constitution in democracy' is better because it provides a focal point for the learner.

The time saved in preparing essay questions (and other subjective-type items) is counterbalanced by the time it may take to read and grade the papers. More important, grading raises problems of standards: one person might think a particular paper is excellent while a second would consider it poor. Care must be taken to clarify just what is to be graded and what points are to be covered in the answer; if emphasis is laid on the understanding of a concept, you may not wish to penalize learners for mis-spelled words or poor grammar. (The learner should at all times know what criteria are to be used.)

Essay questions are probably most useful as a means of providing learners with opportunities to analyse, evaluate or synthesize data, facts or principles in a given situation. The learner must be able to plan his answer and express it in his own words. However, it should be obvious that, where learners' literacy skills are limited, such a format is questionable.

Short answer and completion. These items, as their names suggest, require brief one- or two-word responses: the first, as a response to a question (What is the capital of New York State? _____); and the second, to complete a statement (The city of _____ is the capital of the United Republic of Tanzania). Such items are useful for testing vocabulary, names, dates or the ability to solve arithmetical problems.

This type of test is obviously less subjective than the essay, but it is not as objective as the multiple-choice test, since more than one interpretation is sometimes possible and the person grading often has to make his own decision regarding the correctness or acceptability of an answer.

As with other types of test, certain precautions will help to ensure the accuracy of the short answer/completion format.

1. Make sure that the statement or question focuses on the response you expect. For example: 'The first President of the United States was _____' is presented in such a way that several responses, such as 'George Washington', 'a general', 'a landholder', 'a man', 'rich' or 'elected', are all possible and correct. If you are testing for recall of his name, the item should read: 'The name of the first President of the United States was _____'. This wording elicits the information specifically required.

2. Do not leave too many blank spaces. For example, in ' _____ is necessary along with _____ for plants to _____ ', not only are there several acceptable answers for each blank, but the number of blanks makes it extremely difficult for the student to decide what is required. This item could be re-written in the following manner: 'Three of many essential requirements to ensure plant growth are _____, _____ and _____ ', but grading would still require judgement on the part of the teacher. Would you accept 'fertilizer', 'manure' and 'sunlight' as correct answers? The item could probably be modified further: 'Three essential food requirements for good plant growth are _____ , _____ and _____ .'
3. As you write your items, try to imagine how your learners might interpret them. Ask a colleague to review your test to discover ambiguities. Finally, after using an item, examine learners' responses to see whether difficulties have resulted from the wording of your questions or from failure to clarify a concept during teaching. A test will provide you with feedback not only on group learning but also on your teaching.

True/False. This type of question format permits only one of two responses and is the first of the three objective-type formats to be discussed here. Items are relatively easy to write and can test for specific information, but they are of limited use in measuring depth of understanding. The correct response is chosen in advance and the learner's answer must be either right or wrong.

'Two plus two equals four' is easily recognizable as a true statement. On the other hand, the statement 'Oranges are ready to eat when they are orange-coloured' is both true and false. Some oranges are artificially coloured to enhance sales and some are yellow or green rather than orange when ready to eat. The statement is, therefore, subject to many exceptions and is not suitable for this type of test.

'Meat should be eaten for a balanced diet' might be considered true in cultures where meat-eating is considered essential. However, some cultures prohibit the eating of meat and we know that a vegetarian diet can be balanced, so that it would be preferable to write: 'Most people in European cultures believe that meat should be eaten for a balanced diet'.

The use of *extreme* words such as 'all', 'none', 'never', or 'always', which can provide a clue to the correct response, should be avoided. For example: 'All persons who eat a balanced diet are healthy' is

obviously false and can be recognized as such because it focuses on one word in the question—'all'. Similarly, qualifications such as 'usually', 'sometimes', or 'it is possible' can provide clues that the statement is true.

Since the learner has a 50/50 chance of guessing the correct answer, a score of 50 on a 100-point test could be obtained by sheer guesswork. Limited use of this type of question in combination with other formats is, therefore, strongly recommended.

Matching. In some respects the matching item resembles the completion type mentioned above, but the learner is given a number of plausible answers to select from instead of having to supply his own.

Matching items are used to measure factual information: dates, names, authors, events, etc. Normally two columns are provided: one presents the problem, situation or event, and the other gives a list of possible equivalents. For example:

> On the left (under Column I) you will find a list of fractions. On the right (under Column II) you will find a list of decimal equivalents lettered A–E. For each item under Column I select from Column II the decimal that is equivalent to the fraction. Place the letter of your selection on the line to the left of the numbered item.

Column I	*Column II*
(C) 1. 1/4	A. 0·125
(A) 2. 1/8	B. 0·200
(D) 3. 1/3	C. 0·250
(B) 4. 1/5	D. 0·333
	E. 0·500

When writing matching items, each list should be composed of similar elements, as in the example given above, where all fractions appear in Column I and all decimals in Column II. Responses should also exceed the number of items in Column I; notice that there were four fractions in Column I while Column II had five decimals. Keeping each item relatively short (four to six problems) and writing more than one is also preferable to using one long matching item since it helps to ensure homogeneous columns and is also easier for the learner.

It is possible to write a matching item in such a way that a response option can be used more than once. If you do this, however, make sure that the learners understand what is required of them *before* they begin the test.

Multiple-choice: The multiple-choice test is the most versatile and effective of any discussed. It is useful for measuring a wide variety of learning outcomes, with the possible exception of the ability to organize material, most conveniently tested by means of the essay question. When prepared with care, multiple-choice items provide a highly objective measurement of a learner's knowledge. Their major disadvantage is the amount of preparation involved, but one can gradually build up a stock of items from which to select what one needs.

A multiple-choice item has two parts: the stem, which states the problem (and can vary from a sentence to a paragraph in length), and the options, or possible answers, from which the learner selects the correct answer. For example:

A complete sentence always includes: [stem]

(a) a subject and a verb;
(b) a subject and an adjective;
(c) a verb and a pronoun;
(d) a subject with either a pronoun or an adverb. [options]

There is no hard and fast rule governing the number of options provided, but the four or five normally given should eliminate the possibility of succeeding by guessing.

It is important that all options appear plausible. In the following example, some options are bad:

A balanced diet should always include:

(a) a source of protein;
(b) sugar;
(c) potatoes;
(d) coca-cola;
(e) chewing gum.

Options (d) and (e) are obviously worthless and could serve no useful purpose as potential answers. Nevertheless, if these options were used and (d) or (e) were selected in a large number of cases, the teacher would be forced to consider whether the concept of balanced diet had been properly taught or understood.

When writing an item it is also important to word the stem in such a way as to provide a focus for the correct response. The stem in the following item is unsatisfactory:

Literacy is:

(a) a way to broaden one's life;
(b) a useful communication tool;
(c) a useful tool for further learning;
(d) more important for children.

The stem is obviously incomplete and could be re-written as follows: 'The most important reason for literacy is that it serves as . . .'. Option (d) would also have to be replaced or re-written so as to fit the stem.

TEST CONSTRUCTION

The actual *construction* of a test includes the following three steps:

1. *Item writing.* The items used in testing should accurately reflect the instructional objectives that have been stressed. A test format should be selected in accordance with the level of learning one wishes to assess (for example, rote memorization or the ability to synthesize). The person writing the items must have a knowledge of the subject area being tested.
2. *Item editing.* Once the items have been written, it is useful to ask a second person to review them, to ensure clarity, unambiguity, etc.
3. *Pre-test try-out.* It is strongly advisable to try out a proposed test on a small group similar to the adult participants in a programme. This is a useful means of discovering any possible problems or difficulties connected with understanding or administering the test and estimating the time needed for answering it.

Once the test items have been written and edited they should be set out on the test paper in a clear and easily-read format. Items of the same type should be grouped together, i.e., all true/false items together, all multiple-choice items together, etc. This makes it much easier to give clear instructions to the learners and helps to minimize confusion in responding.

It is also a good idea to arrange items from the easiest to the most difficult. Items that are too difficult at the beginning of a test can lead to discouragement and lack of effort in later sections. The items on each page of the test should be sufficiently spaced out to avoid reading difficulties.

Preparing directions for the learners and the person responsible for administering the tests is vital. The same, or similar, tests will probably be used more than once with the same or different groups and it is

important to administer them in exactly the same way each time they are used. Also, it is only fair to give the learner every chance to display what he has 'truly' learned by making sure he knows what to do.

TEST ADMINISTRATION

It is very possible that the types of test discussed above are unfamiliar to adult students attending basic education or literacy classes and their lack of experience with testing procedures can place them at a distinct disadvantage. For this reason, when administering a test, the instructions prepared ahead of time should first be read aloud carefully and clearly to ensure complete understanding of what is expected. Ample time must be allowed for answering questions related to test procedures and it is always advisable to provide simple practice items for the learners. As we stressed above, a learner's knowledge or understanding cannot be measured accurately if incorrect responses result from misunderstanding of instructions. The learner should be given all possible help to demonstrate his level of attainment.

During testing, it is a good idea to monitor the learners' work to ensure correct procedures are being used and to locate those who may be having special difficulty. At all times remember that adult learners trying to write are making unfamiliar movements with muscles they do not normally use and that they should, therefore, be given special consideration.

The environment in which a test is given may affect the learners' final scores. Ideally, it should be quiet, well-lit and properly ventilated.

Finally, one must note the distinction between *power* tests and *time* tests. In power tests we want to find out only how well an adult can perform and he is allowed any reasonable amount of time. In time tests we want to find out how well an adult can perform within a set period of time and time limits must be strictly adhered to.

Testing groups, institutions and communities

In the course of this chapter we have discussed ways and means of measuring changes in the learning, skills, attitudes and values of individuals. How do these measurement protocols apply to the measurement of changes in groups, institutions and communities?

We must remember that while groups, institutions and communities are, of course, made up of individuals, the properties of a group are not a simple aggregate of the characteristics of the individuals composing it.

When individuals form groups, certain characteristics emerge that belong only to the group as a whole, and not to its separate parts. These *emergent* properties of the group may make its members speak, think and act in a way they would not as individuals. Similarly, while institutions and communities are made up of individuals, they have emergent properties that are not merely the sum of individual characteristics.

Consequently, when we have tested individual characteristics, we have not necessarily tested group, institutional and community characteristics. To do so we will have to conceptualize changes at the group, institutional and community levels. We may wish to measure group cohesiveness among adult learners, the organizational climate of the literacy project office or the level of hope and expectation in a community. We must, therefore, develop indicators to register changes at the group, institution and community levels. These indicators may involve testing of individual perceptions, feelings and learning, but even when dealing with individuals our conceptual sights will be set at new angles.

Summary

In this chapter we have discussed the classical protocols of testing and measurement. The need for indicators of psychological and social conditions was explained and examples of indicators suitable for literacy programmes were presented. Taxonomies of cognitive and affective learning objectives were introduced to sensitize literacy workers to the need to relate testing to teaching. The characteristics of nominal, ordinal, interval and ratio scales were explained. Formats of testing and measurement of specific interest to functional literacy workers, such as questionnaires, interviews, observation schedules and tests were discussed in detail. Reference was also made to the special problems of measuring the emergent characteristics of groups, institutions and communities.

CHAPTER SEVEN

Measurement of change: some rediscovered protocols

In the preceding chapter, we presented what have been called the classical protocols of measurement. They are classical in the sense that evaluators use them when seeking to examine programme experiences objectively. The rediscovered protocols that we shall now discuss do not reject these basic protocols, but their perspective on reality is different and more comprehensive.

The approaches to measurement described previously are seen as objective and scientific. In order to appear objective and scientific, however, they are frequently obliged to parcel reality into manageable portions, selecting a few aspects of behaviour in preference to others because they are unable to measure all aspects simultaneously. In addition, these approaches study behaviour in accordance with pre-established indicators, often dealing only with structured responses. In their attempt to simplify complexity, they lose the immediacy of real life and ignore the dynamics of the situation being studied, so that opportunities for discovering the unanticipated are lost.

In order to overcome these handicaps, evaluators suggest the use of holistic methodologies and measurements, arguing that evaluation should not set out to create universal generalizations about relationships between variables. Since the context of evaluation is often a single project or programme, evaluators must create statements that are educationally significant within that particular context. They should not engage in pure experimentation intended to produce statistically-significant statements at high levels of probability.

Holistic evaluation methodologies and measurement techniques are not as yet clearly defined. Nevertheless, all are based on a common belief that reality should be studied as a whole rather than divided, segmented and truncated to fit designs and measuring techniques. They advocate naturalistic, anthropological, historical and participative approaches to measurement.

Naturalistic approaches

Naturalistic methodologies come to us from the biological sciences. The naturalist is a biologist who works in the field rather than in the laboratory, the most famous being, perhaps, the great Darwin. The naturalist's method is to live among his subjects and study their patterns of life: food-gathering, group life, aggression, co-operation, sleeping and waking, courting and mating, caring for the young, youth, ageing and death. The naturalist observes life as it is lived, looking for patterns and causal links to explain what he sees.

Anthropological approaches

The anthropologist can be thought of as a naturalist who studies human beings. His methodologies are thus naturalistic, but he is interdisciplinary in terms of his analysis. In his search for realistic explanations, he may study bones and skulls; clothing and cooking utensils; language patterns, songs, and folklore; needs and functions; indigenous education systems and political economies.

Anthropological/naturalistic approaches can be used by literacy workers for a variety of in-depth studies, such as the study of the uses to which literacy is being put by adult learners. (Is literacy being used for more effective communication and management, or is it being used merely to increase status?) The reorganization of work and status as the result of new tools and instruments introduced under a functional literacy programme, or the slow change in community norms as the secular encroaches on the sacred and magical, and as new criteria begin to enter decision-making, can also be studied using anthropological approaches.

Historical approaches

The historical method is basically that of the researcher and involves collecting, testing and interpreting data. Often, however, the historian deals with somewhat elusive data about the past—about persons long since dead and gone, about events that happened decades or centuries ago. He cannot readily review evidence or judge the validity of sources and any interpretation of data intended to produce a logical narrative is ultimately intuitive and subjective. Historians have, therefore, developed what they call 'the canons of the historical method',

decreeing, for example, that secondary sources should be used only when no primary sources are available. Historians have also developed careful procedures of *external* and *internal criticism* for use in examining documentation and artefacts. These include grammatical analysis, psychiatric analysis of letters and diaries, carbon-dating and computer analysis.

The literacy evaluator can learn much from these methods, in particular those of the social historian. He can study the documentation generated by the project: instructional materials used, periodical reports written by field workers and administrators, official memoranda, teachers' logbooks, letters and diaries written by adult participants, speeches, photographs, tapes and broadcast scripts. He can go out into the field and see history in the making. He can ask others to recall and recount. He can then synthesize his analyses, looking for patterns and causal links. He can retain convincing explanations, reject unconvincing ones and, finally, test his own hypotheses.

Participatory approaches

Persistent new voices in evaluation claim that change must be planned and evaluated in collaboration with those affected by it. Participatory evaluation would involve collective establishment of objectives and of standards of success and failure. It would also involve collective decisions about what data should be gathered, how and by whom. The analysis and use of data for decision-making would also be a collective enterprise. Participatory evaluation would thus be carried out by the researcher or expert in consultation with programme recipients.

From the foregoing description of the new or rediscovered protocols for the measurement of change three main features emerge:

1. The evaluator must be totally immersed in the environment of the change programme;
2. Holistic methods must be used;
3. Planning and implementing the evaluation of change must be undertaken collectively.

The evaluation methodologies discussed below will be seen to reflect these concerns to varying degrees.

Observation

There are two accepted methods of collecting data on individuals and their lives. We can use a questionnaire or interview to ask about a person's attitudes, skills or actions in certain situations, or we can try to observe his or her behaviour in a situation designed to reveal attitudes and skills. More often than not the old adage 'Actions speak louder than words' rings true.

Observation of behaviour is perhaps the most time-consuming and, some would say, the least objective of data collection methods. However, it can provide us with information about the effects of certain attitudes on behaviour; about the actual *use* of agricultural techniques taught to adults; and about the activities of learners after being taught particular economic or social skills.

Collecting data from real life has a number of disadvantages. Observations are frequently conducted by members of a programme team, whose positive or negative feelings towards the project may 'cloud' their perceptions and introduce bias. On the other hand, observers so far removed from the situation that their knowledge of it is inadequate may make incorrect inferences, thereby introducing bias of another kind. Obviously, we should beware of these two extremes when selecting observers.

Sometimes the presence of an observer may affect the behaviour of the observed. The student, teacher or farmer, knowing that he is under observation, may be on his best behaviour and try to do all the 'right' things while the observer is there. Observers in such cases should try to maintain low profiles.

Of course, if the observer's presence creates problems, or the number of subjects or length of time needed is too great, 'unobtrusive observations' may be used as a source of data. For example, to discover whether writing skills are being used by adult learners, one might go to the appropriate government offices to determine whether the number of people now signing documents such as marriage licences, deeds, passports, etc., is greater than before the programme. To study the change in attitudes towards new farm techniques, one might check the sales records of dealers selling the equipment and supplies needed once those techniques are adopted.

Naturally, in any given situation, an individual can exhibit a multiplicity of behaviours. One can, therefore, imagine the problems that arise when a group is being observed. It is important to determine

in advance the specific behaviours to be observed and to define them in observable terms. For example, if we wish to observe 'co-operation' among members of a group, we might define co-operation as 'getting on with others', and then list some of the behaviours we would expect to see when a learner 'gets on', such as: not speaking out of turn; accepting directions from peers; working well with others in small groups, etc.

Behavioural observations can be recorded in a number of ways, of which the simplest might be to tally each different form of behaviour as it occurs on a pre-designed observation schedule. To note variations in the intensity or quality of behaviour, one might use a scale ranging from 'very much' to 'not at all'. Whatever the method chosen, it should suit the purposes and needs of the situation.

Observational techniques make heavy demands as regards trained observers, time, expense and the need for elaborate field planning. The difficulties of recording data, and of analysing information once it has been obtained, also tend to counteract the usefulness of observation. Nevertheless, an impressionistic use of these techniques in the context of literacy can assist teachers in understanding their learners and provide a means of authenticating information obtained through other, more systematic procedures.

OBSERVATION BY PARTICIPATION

An evaluator may sometimes observe the dynamics of a situation by participating in it. Such observation would normally take place within groups in organizational or community settings. The participant observer may be interested in decision-making processes in a group; group cohesiveness; the organizational health of an institution; leadership patterns in a community, etc. Subjects would generally be unaware that they were being observed.

This approach raises both ethical and technical problems. Is it right to snoop? Perhaps the question when stated in this form is unfair. As participants in social life and work we naturally observe as we participate, and we do our best to evaluate what we observe. It might be better to ask 'What use do we make of what we observe? If the data collected by the participant observer are to be used for *bona fide* evaluation purposes, observation is permissible so long as the privacy of persons involved is protected. Wherever possible, reports should be shown to other participants for comment before finalization and use in the processes of decision-making.

One technical problem involved in observation by participation is related to memory. While a participant observer may occasionally be able to take notes this is not normally possible, since note-taking is liable to arouse suspicion in the observer's fellow participants, thereby affecting their behaviour. Usually, therefore, the observer will have to write his report from memory and, as we know, memory can play tricks.

Another technical problem, that of subjective bias, is more difficult but not impossible to overcome. Today's social scientists have developed techniques that can help the individual to develop perceptiveness and clarity his own value system and the observer to sharpen his analytical skills.

CLASSROOM OBSERVATION

A large part of functional literacy work takes place in group settings—in classrooms, radio forums or demonstration groups—and supervisors or teachers themselves often want to observe group progress. Two techniques are now available to help evaluators *systematize* observation: interaction analysis and cognitive analysis. Together they can be used to develop excellent profiles of group learning and its quality.

Interaction analysis. Interaction analysis, developed by Professor Ned Flanders, makes the basic assumption that interaction between learners in learning situations is good. Interaction should take place not only between teachers and learners but also between learners. In adult groups the need for good interaction for learning and decision-making should be obvious. Professor Flanders has developed a simple chart, accompanied by a set of instructions, for scoring interactions and has suggested ways of interpreting completed charts.

Cognitive analysis. Other educators, while admitting the importance of classroom interaction for good learning, insist that the quality of this interaction must be studied in order to improve learning. They suggest ways in which the evaluator can distinguish between the different categories of quality and content of interactions, record those interactions on paper, and interpret such data for use.

MASS OBSERVATION

Having considered individual and group observation, we must now turn to mass observation. Instead of one or a few evaluators, the whole population may participate in observing the processes, impact and consequences of programmes, so that programme clients themselves

become researchers and evaluators. When studying the impact of literacy programmes on the lives of large communities and recording anticipated and unanticipated consequences of such programmes, mass observation should prove to be an excellent observation strategy.

This technique is associated with the name of C. Madge, an Englishman, who proposed it as a means of studying what he called the 'mind life' of the people. Representative sampling and trained interviewers are not essential; the mass observation technique uses untrained volunteer reporters who make observations in the course of their normal work in their usual environment and describe what they observe fully, clearly and simply. Their reports often include overheard conversations, occasionally supplemented by personal letters or mail received by radio stations and newspapers. Trained interviewers are sometimes used. Partisans of this technique believe that it is easier to detect and account for the bias of an untrained observer than that of a trained social scientist, and claim that mass observation cannot be dismissed as a biased approach to evaluation.

Thematic analysis

How can one organize the immense variety of data collected by means of mass observation or the naturalistic study of the impact of a functional literacy programme on a community?

First, of course, the evaluator must immerse himself in the data. Then he must define the theme or themes touched upon by all those who responded to his questions or provided other data. The process of defining themes has to be heuristic and might raise questions such as these: What issues and events seem to have the highest importance in the community? Do respondents consider any persons or actions responsible for helping or hindering the programme? Are any persons or institutions being blamed? What kinds of fears and hopes are being expressed?

Thematic review and analysis are essential when using naturalistic data. After developing themes, evaluators may have to conduct further tests to control their veracity. Various types and levels of analyses may then be conducted. The direction and pace of change or the administration of the programme may be evaluated. One may ask how wealth is being produced and how it is being distributed. One may analyse the emergent networks of power and status at the community level. One

may even conduct socio-psychological analyses to see whether any late socialization of adult learners is taking place or not.

Content analysis

Another type of analysis of possible benefit to literacy workers is content analysis. Whereas a functional literacy programme might consider the establishment of agricultural co-operatives to be an important part of the agenda for action, the literacy primer and related readings might be promoting rugged individualism and competitiveness. One of the programme objectives might be to teach equality between men and women, but the instructional materials and follow-up books might identify women with traditional—and inferior—social and occupational roles. How can the literacy worker recognize the hidden meanings in the simple stories and statements contained in literacy primers and radio scripts? How can he judge the changes in attitudes and values of people he sees daily simply by studying what they say? How can he determine the cultural biases built into the newspapers and films used in the literacy programme? The answers to such questions can be found by applying the technique of content analysis.

Professor Berel Berelson has defined content analysis as 'a research technique for the objective, systematic and quantitative description of the manifest content of communication'. By specifying 'manifest' content, analysis will be restricted to the syntactic and semantic dimensions of communication. The content analyst is concerned only with what is said. He does not make guesses about the motivations of writer or speaker, nor about possible audience reactions to the communication being analysed.

Content analysis is 'systematic' in two ways. On the one hand, *all* the relevant content must be analysed in terms of *all* the relevant categories, to avoid partial and biased analysis and, on the other, content analysis is designed to establish 'scientific' propositions.

It is also both 'objective' and 'quantitative'. Objectivity, which is enhanced by quantitativeness, stipulates that the categories of analysis should be defined so precisely that different analysts can apply them to the same body of content and obtain the same results.

This type of analysis normally involves the following procedures: (a) defining the body of content; (b) deciding how a sample will be selected; (c) developing a set of categories to be used for classification;

(d) specifying indicators for each category; (e) defining the units of analysis—recording units and context units; (f) training judges; (g) analysing the content of materials; and (h) determining the reliability and validity of conclusions.

Content analysis can be conducted at various levels of sophistication, but the functional literacy evaluator should be satisfied with simple approaches that provide quick but useful feedback.

Process reconstruction

The functional literacy worker is often interested in connections and causes, present tendencies and future possibilities. In such cases, his task involves process reconstruction: what might have happened and what could not have happened? By reconstructing the past, the present may be illuminated and projections about the future may be made.

Viewed from another angle, process reconstruction is employed both by the historian and by the detective. Earlier in this chapter we referred to the historical method. It is worth noting that the detective also uses the methods of the historian, but whereas the latter studies national and international events and communal passions, the detective (whether criminologist or investigative reporter) studies the family, the neighbourhood and individual passions. Although working on a smaller scale, however, the detective is no less scientific in his search for causes. Not only does the modern detective use the science laboratory extensively to identify poisons, teeth, hair, blood and fingerprints, his method is scientific in that it involves fact-finding, hypothesizing and testing hypotheses by way of process reconstruction. The detective's ultimate question is: How could it have happened and how could it not have happened given a particular pattern?

Professor Michael Scriven, whose work has been referred to earlier, has tried to demonstrate the logic of the process reconstruction technique, which he calls the 'modus operandi'.

Modus operandi

Modus operandi literally means the manner of operating. As Professor Scriven has pointed out, the modus operandi of a particular cause is always 'an associated configuration of events, processes or properties usually in time sequence, which can often be described as the *characteristic causal chain* (or certain distinctive features of the chain)

connecting the cause with the effect'. It is thus the method of pattern recognition and, consequently, that of the detective, the historian, the anthropologist and the engineering troubleshooter.

The modus operandi methodology suggests that the world of the evaluator should be ordered in terms of the phenomenon or event that has occurred (X), its possible causes ($A_1, A_2, A_3, A_4, \ldots, A_{n-1}, A_n$) and the circumstances surrounding the cause-effect chain (C).

The evaluator is primarily interested in the study of causes and effects *in particular contexts* and *not* in generalization across contexts. Thus, while C ought to be described, it may be taken for granted and need not be treated as a variable.

X should also be described. What effect is being studied? Is it the appearance of a new variable? Is it the increment or the decrement of an existing condition? Is it the appearance of a new configuration of variables? The evaluator must always be prepared for unexpected consequences and side-effects. In fact, goal-free evaluation, also advocated by Professor Scriven, lends itself admirably to the modus operandi procedure.

The essence of this methodology is what Professor Scriven calls 'cause-hunting'. A list of probable causes that could, singly or together, have produced the given effect must be developed. This implies, of course, that the evaluator has considerable background knowledge, but cause-hunting need not devolve on him alone; it should provide an excellent opportunity for everyone involved in the project to participate in the evaluation process.

This 'quasi-exhaustive' list of causes may be generated by theory, analogy or on the basis of empirical knowledge. More probable causes can be added at a later stage when the list is tested for completeness. Again, the causes in the original list need only make 'claims to weak local determinism'. It is sufficient to suppose that Xs observed or to be explored are *usually caused* by these As in given Cs.

The stages in applying the modus operandi (MO) methodology are as follows:

1. Check for the presence of each A. If there is only one, that is the cause.
2. If more than one A is present, check for complete MOs. If none, then none of those As was a cause.
3. If only one MO is complete, the A with which the MO is associated is the cause.

4. If more than one complete MO is present, the associated factors are co-causes.

An MO analysis of different co-causes will reveal the causal links between a particular set of causes and their effects. 'Tracers' can be introduced to provide information about routes and networks. By conducting MO analyses in more than one situation, control group arrangements can be imposed on the methodology. The MO methodology is thus a useful device in situations where experimental and quasi-experimental designs cannot be employed. Where those designs can be employed, it can be used to supplement them.

PARTICIPATIVE EVALUATION

In Chapter Three on pages 22–24, we introduced the concept of participative evaluation and described the various steps involved in conducting participative research and evaluation. Various groups are currently displaying considerable interest in this type of evaluation, but so far they have made little methodological progress.

CASE STUDIES

Case study investigations focus on a single case or a group of similar or contrasting cases. The 'case' may often be defined to suit the interests of the evaluator. For instance, the functional literacy worker might select as a case Cuba's eradication of illiteracy, aspects of the mobilization for literacy in Maharashtra State, India, the writing of a literacy primer or the effect of literacy on the lives of half a dozen neo-literate farmers.

No single methodology can be applied to all case studies. A holistic approach is essential and any effective technique for collecting data or testing hunches and hypotheses can be employed. It is usually asserted that one cannot generalize from case studies, but this is not necessarily true; a good case study may present the essence of experience that is clearly transferable to other situations and settings. The occasional assertion that case studies are useful in the infancy of a discipline, but are of little value once that discipline has matured is equally indefensible; a good case study may prove helpful in any discipline at any stage.

The place of holistic techniques in evaluation

It should be understood that not all evaluation questions can be best answered by applying the holistic methodologies and techniques we

have just outlined. Much evaluation work will involve classical techniques such as sampling, subject matching, use of control groups and structured testing, which will be discussed in Chapter Eleven. Even when using naturalistic and historical approaches, the evaluator will often seek and use hard data collected for other purposes in other contexts.

Holistic techniques are most useful in answering the following types of question, of critical importance to many projects.

1. How does literacy change the lives of new-literates? Do new-literates use different strategies for coping with the world around them? Do they acquire new identities in the process of becoming literate? Do they appear different to the people round about them? Have their relationships with local institutions changed?
2. Is the opportunity to perform the role of a literacy teacher changing the life of that teacher? Is the literacy teacher becoming a new secular leader in a formerly traditional community? What effect does this have on the existing networks of status and power?
3. How does the illiterate cope with his life in the community? What is his self-concept? Does he really feel inferior and helpless?
4. What uses are being made of literacy skills by new-literates?
5. If productivity has increased, and with it the farmer's income, where is that income going? How are people actually spending their money? Do they wish that they could spend their money differently? On what? Do they wish that they could save it, and through what saving and banking institutions?
6. Is any information which is necessary for development failing to reach communities? Is any becoming distorted? Is it possible that some information should be disseminated by the media? What information must be transferred in interpersonal settings?
7. In a functional literacy project, are literacy and production truly integrated? What does integration mean from the point of view of the people whom the programme is intended to help?
8. Is the newly-established co-operative fulfilling its goals? What is its impact on the lives of participants? Has it produced any co-operative fall-out in other sectors of life in the community?
9. Have the organization and arrangements for teaching literacy to illiterate adults taken hold as semi-permanent pattern? What does institutionalization mean in this case?

10. How is decentralization working? How local is local initiative in programme planning and implementation?
11. If recruitment in a programme was satisfactory, why did the programme eventually fail?
12. How have participants reacted to political commitments shown by national leaders and programme commitments shown by field staff? How widely are these reactions shared?
13. How did that unfortunate rumour about family planning spread in the community?
14. What is the impact of the programme as a whole on the community as a whole? What did the programme do to the existing power structure? Did it create a sense of movement; a reason for hope?

What is holistic research?

We have discussed various types of holistic research, including the observation and modus operandi techniques. In practice, we use some or all of these techniques in accordance with our needs. Nevertheless, the evaluator should always go into the field to observe and interact with those whose lives the programme is supposed to be changing. He must be empathetic and perceptive. He should understand, but not attribute, motives. He must observe people at work in formal and non-formal settings. He should read and examine what others have written, and learn from the records they have left behind. To old questions he will add new ones, developing themes that are part of the consciousness of teachers, administrators and participants. He will analyse these themes, make lists of hypotheses about causes and effects, look for obvious inferences and develop conclusions.

Holistic evaluation is not entirely subjective. The evaluator may systematize his observations by using observation schedules. In making judgements, he may make use of non-parametric statistics to test the reliability of his evidence. In order to eliminate the possibility of personal bias, he may want to work in a team, with all members trying to find answers to the same questions.

Self-evaluation

A literacy programme evaluator can ask his subjects—adults, teachers, supervisors and administrators—to engage in self-evaluation. How much do they think they have learned? What changes do they notice in

their attitudes? Do they think they are doing the best job possible? What kind of further training do they require?

Depending on the questions asked, how, and in what context, self-evaluation can provide excellent evaluation data. Sometimes, no doubt, when adults are questioned directly they reply politely but not correctly, making socially acceptable responses and inventing answers to questions about children, eating habits and family income. But once rapport is established between the questioner and the adult subject and the latter understands why information is needed and to what uses it will be put, self-evaluation can prove both reliable and useful.

Self-evaluation requires both capacity and courage. The individual should know what can be achieved. He should realize that different people and groups have different norms and standards. He should have awareness that allows him to engage in dialogue with himself. He should also have the courage to express the results of self-evaluation. He must be neither insecure nor timorous.

Both the capacity and the courage needed for self-evaluation can be taught and learned. Indeed, this ought to be one of the teaching-learning objectives for all functional literacy projects. Today, possible techniques for consciousness-raising are widely debated, but discussion of them lies beyond the scope of the present monograph.

Summary

In this chapter we have discussed the new and rediscovered protocols of measurement. These do not reject the classical protocols but go beyond them. Variously called holistic, naturalistic, anthropological and historical, these approaches display one or more of the following features: (a) the evaluator's immersion in the environment of the programme to be evaluated; (b) holistic approaches to measurement; and (c) participative planning and implementation of evaluation.

Holistic techniques were shown to include observation of various kinds; thematic and content analysis; process reconstruction and modus operandi procedures. Participative evaluation and case study techniques were also mentioned. Finally, the immense possibilities of self-evaluation, both in the learning process and as a source of objective information, were examined.

CHAPTER EIGHT

Implementation data for management and evaluation

All educational and development programmes leave footprints behind them. Careful examination of those footprints can give programme specialists a considerable amount of information: How many adult men and women were enrolled in literacy classes? How many dropped out, and when? What was the turnover of literacy teachers in the programme? How many visits to the field did supervisors actually make? How much did it cost to produce one copy of a literacy primer? How many adults bought the primer? How many did not?

In the normal course of their implementation, programmes can generate data that, if properly stored, can be put to good use by programme specialists and evaluators. Indeed, with some systematization, implementation data can constitute a valuable management information system (MIS).

An MIS may be characterized as follows:

1. It must supply relevant and accurate information. It must not collect more information than will actually be used. It must collect the information that decision-makers will need at every stage of programme development.
2. It must be economical and reliable. Collecting information costs money. The MIS should be matched with the project's budget.
3. It must produce information for users *when they need it.* Information that arrives too late is of no value; it is not worth collecting.

In any large-scale functional literacy project, an MIS is essential. It is the only means of synchronizing decision-making at the national, regional and local levels, while simultaneously decentralizing it. The kinds of information that should go into an MIS are (a) contextual information or base-line data; (b) data providing profiles of a project at various points in time; and (c) data on the effects of the project on individuals and communities.

A word of caution

Before entering into a discussion of the kinds of implementation data that could be collected and the variety of registers, forms and tabulations that could be used to record and store such data, a word of caution is in order. The list given here of possible data that could be collected is comprehensive. However, it should be understood that one does not always need to collect all types of data. What is possible is not always feasible. As pointed out in the Preface, this monograph seeks to inform, to develop complete descriptions, not final prescriptions. As a training monograph, it must detail all the options available to the functional literacy evaluator. Individual evaluators however, may concentrate on only a few aspects of a project and choose to collect only some of the implementation data it generates. It would be a distortion of priorities for a small project working with scarce resources to collect exhaustive data. The collection of unnecessary data unlikely to be used in decision-making is a waste of resources. Each project, therefore, must decide upon a minimum set of needed implementation data. This is a matter of policy that must be decided before any collection of implementation data for management and evaluation is planned.

Bins, bowls and baskets for implementation data

The bins, bowls and baskets used in gathering implementation data in a literacy project are cards, forms, tables, schedules, registers, logbooks and workbooks.

Specific data records include: (a) individual participant profile cards; (b) literacy teacher profile cards; (c) forum leader profile cards; (d) extension worker profile cards; (e) supervisor profile cards; (f) attendance registers and class profile tables; (g) workbook summary tables; (h) teachers' logbooks; (i) discussion forum logbooks; (j) agricultural demonstration logbooks; (k) family survey forms; (1) survey forms for farms, factories or other production units; (m) institutional survey forms; (n) community survey forms; (o) field visit report forms; (p) monthly/quarterly/yearly report forms.

Let us stress that not all projects need collect such elaborate data. Each literacy project must decide what implementation data it needs. Some may be numerical; some may be descriptive, anecdotal or judgemental. Most projects will make use of attendance registers,

teachers' logbooks, supervisors' report forms, field observation schedules, workbook summary forms, progress cards for students and teachers, standardized periodical reports, etc.

Whatever the size and scope of the MIS, it will entail some stationery and printing costs, besides making demands on the time of teachers, supervisors and administrators. However, this expenditure may be as low as 5 to 10 per cent of the cost of collecting and systematizing data in an independent evaluation effort and its educational dividends are truly enormous. Forms and tabulations can be designed to provide feedback to the teacher or supervisor while he completes them, enabling him to take immediate steps to improve his part of the programme. This is extremely important, as evaluation studies are frequently criticized on the grounds that feedback often becomes available only when it is already too late to make any use of it. Collection and collation of data will point the need for visits to exceptional field sites, for case studies and for systematic analysis of causes and effects.

INDIVIDUAL PARTICIPANT PROFILE CARD

It is difficult to conceive of a literacy programme with no interest in the performance of its individual participants. In a functional literacy programme, this performance has to be judged not only in terms of reading, writing and arithmetic, but also in terms of attitudes learned and new occupational and economic skills acquired. Data on an individual's use of new attitudes and skills may also be collected, as exemplified in the following clusters. This does not mean, however, that each cluster can be dealt with in a single test or observation schedule.

Cluster I
Speech; reading; writing; computing; picture literacy.

Cluster II
Economic knowledge (vocational); health knowledge (scientific); family-life knowledge (scientific); civics knowledge (social scientific); other.

Cluster III
Use of: reading; writing/expression; arithmetic; graphic/pictorial skills.

Cluster IV
Adoption of: recommended economic practices; recommended health practices; recommended family-life practices.

Cluster V
Participation in: economic institutions; social institutions; political institutions; religious institutions; traditional institutions.

Cluster VI
Effects of information on: radio listening habits; written materials; the sending of children to school.

Cluster VII
Modernity scale; empathy; motivational scores; n-achievement; local/cosmopolite scores.

Cluster VIII
Cognitive skills; reality versus ceremonial orientation; cause/effect stance.

Cluster IX
Acquisitions: radio, bicycle, sewing machine, etc.

Such data can be collected in time series and may be used to establish correlations between the following: (a) age and literacy achievement; (b) age and adoption of new practices; (c) sex and change-proneness in particular aspects; (d) residence and modernity; (e) occupational needs as regards literacy and literacy retention; (f) duration of inputs and performance; (g) intensive versus non-intensive input and behavioural change; (h) land ownership and adoption of agricultural practices; (i) ownership of radio, bicycle or sewing machine and over-all modernity. Needless to say, these data can be used for both diagnosing and evaluating.

THE INDIVIDUAL IN THE PRODUCTION SETTING

The above profile will provide a considerable amount of information about the individual as a learner of literacy skills, an absorber of new attitudes and a master of improved economic skills. However, knowledge of improved skills and acquisition of better attitudes is not enough; to study an individual as a producer, we must study him in the

setting in which he is productive. The particular constraints and facilities of the setting in which economic activity takes place will affect production. This means that functional literacy workers will need to collect periodical data on, for example, an individual peasant's farm or the factory in which a worker is employed.

Data on farm size, animal husbandry, farm management and reinvestment can provide useful information about the innovative behaviour of the farmer. In the factory setting, changes in productivity, income and safety habits can be explored to advantage.

PROFILES OF CHANGE AGENTS

Change agents are seldom introspective. They rarely want to observe themselves or question their own abilities to implement plans in the field. Thus, while most literacy projects collect data on their clients—the adult men and women in their programmes—they seldom study their teachers, forum leaders, extension workers and supervisors. Nevertheless, the training, background, social origins, role perceptions and work patterns of the change agents themselves can explain many things that happened and many others that did not.

The role of the *literacy teacher* is all-important. Teacher data cards might record demographic information on the teacher, his previous teaching experience, data on his selection, training and service within the programme.

In some literacy projects, adults meet in a discussion forum, a radio forum, a film forum, or a combination of these, in which audio-visual material (radio broadcasts, films, flashcards or flannelgraphs) is used to initiate and guide discussion on a topic of interest to adults. The *forum leader* is sometimes the literacy teacher himself, but not always. In either case, it might be useful to develop a data card on the forum leader in order to relate impact on adults to leader competency and characteristics.

Data on the *extension worker* might also be collected and stored. Here again, it must be emphasized that a knowledge of the extension worker's education, background and previous assignments can throw considerable light on the successes and failures of the programme.

In small projects, *literacy supervisors* may be few in number and individually known to the project staff. In such cases, it is probably unnecessary to collect data on them formally. In larger projects, where supervisors are more numerous, it becomes essential to collect data on them.

Compiling data on an individual produces considerable anxiety in the person concerned. When a data card is designed to include information related to performance, supervisors may think that 'big brother' is out to get them. Supervisory staff and extension workers in rural areas generally enjoy considerable freedom and develop highly individualized patterns of work. When data is compiled on their work patterns and performance, they can become extremely alarmed. It is therefore necessary for project directors to discuss the objectives of collecting such data with supervisors. There is no short cut to open and honest communication. It must be stated that the main concern is to relate supervisors' background and characteristics to various aspects of the programme and not to exercise control, although, in all honesty, data collection does have a disciplinary aspect as well.

CLASS RECORDS

The *attendance register* is the basic class record. It should show the names of those enrolled in a class, their age, sex and attendance. Reasons for absence can often also be displayed in this register. A monthly summary should be included, designed to provide the teacher completing it with instant feedback on the functioning of the class.

A *class card* is sometimes useful as a source of information about instructional materials used, outside inputs into the class, time utilization patterns and attendance and drop-out rates.

Adult learners in functional literacy projects quite often use workbooks and partially-printed exercise books to practise their reading and writing skills. *Workbook summaries* can provide excellent evaluative feedback on the progress of individual learners, as well as on that of the group of learners as a whole. Such information can be used to distinguish between slow learners and good learners; to discover patterns of errors to be taken into account when revising materials; to evaluate teacher effectiveness and a host of other things. The use of the workbook for evaluative purposes is discussed in Sohan Singh's monograph, *Learning to read and reading to learn: an approach to a system of literacy instruction*, published in this series of monographs (see list at front of book).

A literacy teacher may be encouraged to keep a *logbook*, a bound exercise book in which to note, on a daily basis, important classroom events and their possible significance as well as his view of adult learners' progress, feelings and relationships with him and with their fellow-participants. Why does he think he can 'get through' to adults

on one day and not on another? What does he intend to do about it? It must be pointed out to the teacher that the logbook should not reduplicate information that will be included in other records and tabulations. It should be used to record aspects of the inner life of the classroom that evade quantitative measurement. The teacher must learn to remember important events as well as to store insights and use them to increase his perceptivity both as a teacher and as a person.

The *discussion forum* must also be studied. The above-mentioned types of class record can be used to collect information on a discussion forum but, in some cases, a logbook might be more appropriate. It should include information about: (a) where the session was held; (b) the number of adults present; (c) how many of those present belonged to the literacy class; (d) the topic discussed; (e) materials used for discussion; (f) questions that remained unanswered and to which answers must be found by the extension worker or a specialist; and (g) general remarks on interest and participation.

Agricultural demonstrations and other inputs

We have repeatedly stated that a functional literacy project is not merely a literacy project, but a social change project in which the teaching of economic/occupational skills is the basic objective. It should follow that the inputs made into the programme are not only the teaching of reading, writing and arithmetic, but also that of economic skills, marketing or agricultural demonstrations. In addition to instruction, there will be material inputs such as fertilizers, new tools, high-yield varieties of seeds, improved equipment, etc. A record must be kept of all these inputs, demonstrations organized, and materials supplied to individuals, classes and the community as a whole.

FAMILY SURVEY

Another important unit indicating the impact of a functional literacy programme is the family. Since a functional literacy programme is meant to promote socio-economic change, the adult learner's *performance* in the family must change. Every individual programme will have to determine its own objectives with regard to desired changes in the family; these may be related to general cleanliness, family diet, child rearing, health habits, the sending of children to school, family consumption patterns and family decision-making.

INSTITUTIONAL SURVEY

In some functional literacy programmes the creation of new institutions may be an important objective. In one rural sector project, the establishment of an agricultural co-operative may be the essential feature. In another, it may be the setting-up of an institution for self-government or local decision-making. Generating and maintaining sufficient information on the creation and subsequent growth of such institutions is important. A project's co-operative specialist should be able to develop records to document the establishment and growth of a co-operative. Similarly, the persons concerned with the establishment of a 'group of elders', a *panchayat*[1], or a local development committee, should be able to generate and store information on the development of their particular institution.

COMMUNITY SURVEY

A synoptic view of the community in which a functional literacy project is being carried out may be desirable. Well-planned social change projects do not end up by helping a restricted number of individuals within a community, even though they may begin working with just a few people. A generative spread effect is almost always implicit in the design and carrying out of such projects. A community survey is thus doubly important. While showing whether benefits are spreading into the community—as has been implicitly or explicitly assumed—it should guard against the possibility that the functional literacy project, by working only with selected individuals, may be perpetuating and strengthening existing privileges.

Community surveys may turn into elaborate and comprehensive base-line surveys (at the beginning of a project) or terminal long-term evaluation (at the end of a project or one of its phases). However, short local community surveys can be planned for each village or group of villages where a programme has been started. The following information may be useful: types of home; drinking water resources; new enterprises; new institutions (bank, co-operative, library); and the growth of infrastructure.

FIELD INSPECTION REPORTS

Supervisors are not the only people who visit literacy classes. Programme officials from the project office regularly visit classes to see

1. A *panchayat* is the unit of local self-government at the village level in India.

literacy work at first hand and test the feasibility of their ideas in the field. However, it is not unknown for each individual visitor from the project office to keep silent about what he or she learns from these field visits, and officials often repeatedly visit the same classes—those that are close to the road or fit conveniently into their time schedule.

A standardized field inspection report, to be filed in the project office, with open access, can serve several useful purposes:

1. The process of filling in the inspection report form forces the field visitor to systematize observation, think about what he has seen and look for themes that seem to determine motivations, processes and actions;
2. Information gathered by one member of the project team can be shared with the rest of the team;
3. A more representative sampling of class sites can be assured.

PERIODICAL REPORTS

Projects and programmes all over the world publish periodical reports—monthly, quarterly, yearly and mid-term. Unfortunately, periodical reports are often nothing but elaborate exercises in forgery. They seldom, if ever, offer a balance sheet: credits are shown, debits are hidden. Nothing that goes wrong is ever reported: no bad decisions, no conflicts in the organization, no breakdowns, no wastages ever find a place in the monthly and quarterly reports.

These reports, however, are potentially instructive group evaluations. In this political world, the typical periodical report submitted to ministries and funding agencies may continue to serve important purposes. At the same time, separate evaluative periodical reports could be produced for the benefit of project teams, even if they cause discomfort and insecurity.

Mechanical aids for data analysis

The collection of implementation data is often easier than its analysis. Data recorded in attendance registers, cards, logbooks, tabulations and reports has to be analysed. It has, therefore, to be organized, processed and manipulated in various ways. When the figures involved in literacy evaluation reach the thousands (say, 3,000 teachers or 75,000 adult learners), the problems of data analysis become extremely complex and difficult.

We have all heard of the computer, that wonder of technology. Computers have fantastic memories and can process the numbers fed into them at the speed of lightning. They can accomplish in a matter of minutes what would take a professor with pencil and paper a lifetime. These marvels are no longer to be found only in highly technological societies. Today, there is hardly a Third World country that does not have a computer at work within government or business, or in one of its universities. Literacy projects dealing with large amounts of data may wish to avail themselves of computer facilities where these exist. If computer use is envisaged, a computer specialist should be consulted at an early stage so that evaluation instruments can be designed for computerized data analysis.

Between pencil-and-paper and computer analysis lies an intermediate technology involving the key-punch machine and the card-sorter. The key-punch machine functions like a typewriter and can be used to record both alphabetical and numerical data on a standard computer card. The computer card has 12 rows and 80 columns and thus provides 960 positions for punch holes. If necessary, more than one card can be used to record data on a single individual or social unit—such as a community. Key-punching is relatively speedy; cards are easy to handle and store; and, most important, the data on the cards is presented in a form convenient for analysis.

The card-sorter, as its name suggests, sorts the computer cards carrying data. It can be set to separate cards with any given pattern of key punches. The sorting process is carried out one column at a time, but the speed of a typical sorter is reasonably high and as many as 39,000 sorts can be made per hour. The card-sorter is compact, efficient and quite inexpensive.

Readers are advised to contact a local university or the International Institute for Adult Literacy Methods for further information about books and materials to assist them in coding, preparing code sheets and punch cards, and using the sorter for data analysis.

Summary

This chapter has discussed the need for the systematic organization and storage of the data automatically generated during the implementation of a literacy programme. It is pointed out that such data can be collected with very little additional effort and expense if a project has made prior plans to perform this task. The data collected may relate to

individuals or groups, to processes or contexts. Time series of data can also be generated. Implementation data can provide a management information system (MIS) in addition to answering typical evaluation questions. Reference has been made to the mechanical aids available for the analysis of such data.

CHAPTER NINE

Special evaluation tasks for functional literacy workers

In the previous chapter, we discussed the collection, organization and storage, for evaluation purposes, of data produced during the carrying out of a programme. We will now suggest means of assembling data for special evaluation tasks in a functional literacy programme.

As we have stressed, the formulation of an evaluation policy *must precede* the development of an evaluation programme. Literacy workers must select one particular set of evaluation questions from the vast array of possible issues. Since each functional literacy programme has its own socio-political setting, it will naturally have its own evaluation priorities.

This chapter will be devoted to a number of *evaluation tasks* commonly undertaken by functional literacy projects. Not all functional literacy projects will, of course, wish to perform all the tasks listed below, and some may be interested in types of evaluation to which no reference has been made. Nevertheless, we shall comment briefly on:

Base-line surveys;
Needs assessment and needs negotiation;
Achievement testing (of reading, writing, arithmetic, occupational knowledge and other skills);
Formative evaluation of instructional materials (graphics, primers and readers);
Measurement of changes in attitudes;
Evaluation of teacher effectiveness;
Evaluation of systems of action;
Cost-benefit and cost-effectiveness analyses;
Community surveys for the study of impact.

We shall define each of these evaluation tasks in turn, stating its objectives, suggesting how it could be carried out and listing the uses

to which the information it generates could be put. Specimens of evaluation tools that have been used in different projects all over the world have been assembled by the author in a *Dossier of instruments*, available on request from the International Institute for Adult Literacy Methods, Tehran, Iran.

Base-line surveys

Most functional literacy programmes conduct a base-line survey. If not, they often regret it later on. A base-line survey, as its name suggests, sets out to establish a social, economic and cultural base-line against which programme effects can be measured at different points of time during the life of the programme. Functional literacy programmes normally expect to make more men and women literate. They also hope to bring about some attitude changes among adult learners. They wish to improve health practices within communities. They may also have large-scale objectives, such as improving agricultural productivity, establishing co-operatives, creating marketing and storage facilities, developing communal harmony, and encouraging social and political participation. The processes involved in fulfilling these objectives are teaching, animating, organizing, communicating and building new institutions. Eventually, functional literacy workers have to ask themselves whether their teaching, communicating and organizing made any difference. Did the communities, and individuals and groups in those communities really change?

In order to answer such a question, functional literacy workers need a profile of the community *before* their programme was introduced. They also need a profile of the community *after* their interventions. Only then can they assert that their work in the community has produced changes. This is the main reason for conducting base-line surveys.

A base-line survey can be limited or lengthy, depending upon the objectives of the functional literacy programme. If these are modest, the base-line survey should also be designed on a modest scale. If programme objectives are comprehensive and aspirations for change very high, a much fuller base-line survey should be planned.

Some important points must be borne in mind. First, since data collection costs money, only data that will actually be used should be collected. On the other hand, returning to the field to collect information that should have been gathered the first time round is even more

expensive, so that data that is sure to be needed later on had better be collected at the start. Secondly, base-line surveys are most useful when programme specialists know in advance what type of information they expect the data to provide and what claims they intend to make for the impact of their programme. Thirdly, base-line surveys must have theoretical sophistication. They must illustrate links between causes and effects and demonstrate relationships between variables. Fourthly, the information collected through base-line surveys should be easy to use either as a whole or in part. The production of a base-line survey that is never used is a regrettable—and, unfortunately, not infrequent—occurrence.

Base-line surveys may use a wide variety of measuring formats and instruments, including numerical data, attitudinal questions and observation inventories. A specimen of a base-line survey instrument, used in a functional literacy project sponsored by the Unesco Experimental World Literacy Programme, is included in the *Dossier of instruments* referred to above.

Needs assessment and needs negotiation

Everywhere the organizers of social action programmes are being challenged to meet the *real* needs of the people they seek to serve. Functional literacy workers are no exception. Understandably, therefore, most functional literacy programmes start by defining client needs. This process, called 'needs assessment', is often undertaken as part of an initial base-line survey.

It is quite usual to ask the prospective adult learners themselves what they feel their needs to be. For example, they might be asked: What are your needs as regards agriculture, health, home and daily life? What measures do you think are called for to meet those needs? What kinds of programmes should be set up to improve social and economic conditions in your community? Answers to such questions are certainly helpful to decision-makers, but assessment of needs is not as simple as this.

In a sense, asking people to state the needs they feel is unfair and possibly even dishonest. Social action programmes, after all, do not merely seek to fulfil *felt* needs. They are, more significantly, designed to fashion *new* needs. They are often intended to make individuals want better health, balanced diets and planned families; to make groups eschew conflict and acquire new collective identities; and to

make institutions assume new missions and undergo renewals. All these objectives entail the learning of new needs.

Nevertheless, functional literacy workers have a special interest in felt needs. In the first place, some of a community's felt needs may be both genuine and urgent, and fulfilling them may be the best way of gaining a foothold in that community. In the second place, functional literacy workers can build on felt needs in order to fashion new ones. The new must be stated in terms of the old; fashioned needs must be related to felt needs in order to tap existing motivational structures.

Professors Richard T. Coffing and Thomas E. Hutchison have rightly suggested that the assessment of needs should be conducted in terms of 'who needs what, as defined by whom'. The clientèle should be identified first. Client needs must then be stated separately by a number of definers to obtain different sets of statements of needs. For example:

1. Needs of farmers as defined by the farmers themselves.
2. Needs of farmers as defined by local extension workers.
3. Needs of farmers as defined by managers of agricultural marketing co-operatives.
4. Needs of farmers as defined by decision-makers in the project itself.
5. Needs of farmers as seen by the national planning board.

Other clients may then be identified—for example, adult women or school leavers in the community. Their needs may also be variously defined. In this way, different profiles of needs may be developed for one and the same community.

Assessment of actual needs may be carried out by means of individual interviews or questionnaires. Distinctions should be drawn between definers who are individuals and definers who are groups. Large groups may be sampled if it proves too difficult to interview or send questionnaires to all members.

NEEDS NEGOTIATION

The process outlined above will provide a number of different profiles of needs. How can one combine these profiles? How can one assign priorities among needs? How can one develop a final profile of needs on which to base a programme of action?

Some might want to leave these problems to the decision-makers, who should have obtained sufficient information to be able to make

the right decisions. Recently, however, those in favour of participative approaches to planning and evaluation have suggested that building the final profile of needs is too important to be the monopoly of programme specialists; those whose lives are to be changed must participate in it.

We are in complete sympathy with this view. People are unlikely to learn to take part unless opportunities for participation are provided. They are unlikely to ask to participate unless they have already been allowed to do so. One method of developing a final profile of needs would be for decision-makers and client groups to sit down together and, on the basis of information available to them, produce a final assessment of needs through a process of negotiation. Once power has been tasted by currently powerless adult groups, they will probably learn how to exercise it. The process of needs negotiation and participation will then become genuine.

Achievement testing

An important consideration in any functional literacy project is the measurement of adult achievement in reading, writing, arithmetic, economic knowledge and skills taught as part of the programme of instruction. We have already mentioned achievement tests and the taxonomies of educational objectives, stressing the need to test the achievement of the intended objectives. We shall now make a few additional comments on the design of achievement tests.

ORAL READING TESTS

In the early stages of teaching adults to read, most reading tests will have to be oral. Before deciding to use this type of test, one must determine the criteria for an acceptable reading of a set passage, or one picked at random from a small selection of passages. These criteria should be pre-tested for a common standard between different assessors. Oral reading tests may sometimes include recognition of letters, syllables and words by adult learners.

WRITTEN TESTS OF READING

Written tests of reading tend to test vocabulary level, comprehension and application of ideas. Since such tests must provide a response mechanism they are generally of the objective type. A few specimens are included in the *Dossier of instruments*.

TESTING WRITING ABILITY

At the start of literacy training, learning to write may involve much copying of model words, phrases and sentences. Later on, learners may be required to take dictation and write words, phrases and sentences. When preparing writing tests, the level of writing ability to be tested must be clearly defined. The criteria for good writing must also be established and agreed upon.

ARITHMETIC TESTS

Since arithmetic is a highly structured subject, in which the 'right' and the 'wrong' are universally recognized, arithmetic tests are comparatively simple to design. Furthermore, the various test items in an arithmetic test can be presented in increasing order of difficulty and complexity.

Teaching arithmetic to adults should not be restricted to the mastery of computational skills, but should include knowledge of how to use those skills in different social settings. In other words, adults should be taught how to use arithmetic in farm management, or in buying and selling. This is one area in which it may be possible to borrow across projects. For examples of arithmetic tests see the *Dossier of instruments*.

TESTING OCCUPATIONAL KNOWLEDGE

Functional literacy programmes are frequently designed for particular occupational groups (fishermen, textile workers, wheat farmers, etc.) and are built around specific economic themes. Teaching improved productivity and management skills is often the first objective of the programme. Participants' learning of the subject-area—for example, cotton farming, wool carding, fishing, housekeeping—must, therefore, be tested.

In testing for occupational learning, functional literacy workers must distinguish between the acquisition of information by adult learners and the application of this information in their homes, farms or factories (see the taxonomies of educational objectives in Chapter Six). If a farmer knows how much space should be left between rows of cotton plants, but does not apply this knowledge in his own field, his learning has not been of much use. On the other hand, if he has not absorbed information, he will be unable to apply it! In testing occupational knowledge, functional literacy workers must use a combination of achievement tests (to test the learning of information) and

observation (to see whether the information learned is being used in the life and work of participants).

TESTING LEARNING OF SKILLS

Testing the newly-acquired skills of adult learners is another difficult task. The person testing must be qualified to judge how far the performance of the skill he is observing approximates to the ideal. Skills may be tested by the results they produce in terms of productivity, safety and avoidance of wastage. In all cases, the standards for acceptable performance and criteria for judgement must be laid down in advance.

Formative evaluation of instructional materials

One of the most frequent evaluation tasks in functional literacy programmes involves the testing of instructional materials. Such testing may be carried out during the initial design, or formation, of instructional materials and is sometimes called formative evaluation. Materials may be produced in limited editions and tested in a pilot project to provide an opportunity for thorough revision before general distribution.

Functional literacy projects normally deal with materials that are graphic, verbal, or both. So far, although the radio has been in existence for decades and the cassette recorder for some years, audio and audio-visual approaches to teaching have not yet made any great impression on the world of literacy.

DID THE PICTURE CARRY THE MESSAGE?

Literacy workers often test the illustrations used in primers, flannelgraphs, flashcards or posters. Such testing rests on these assumptions: (a) some cultures are graphically-oriented and some are oral; (b) cultures have their own peculiar sense of perspective; and (c) most illiterates need to learn to 'read' pictures, that is, to acquire visual literacy.

The developers of graphic materials want to know whether learners are able to recognize what the artist set out to depict. Simply showing such material to adult learners and then questioning them about the different elements in a picture and their meanings can be highly instructive. More ambitious evaluation projects may seek information about: (a) the continuity and flow of graphics in a set of flashcards; (b)

the use of different colours, graphic techniques and conventions in visual design; and (c) the relative contribution of pictures and their captions to the communication of messages. Examples of approaches and instruments used in this kind of evaluation are included in the *Dossier of instruments*. Readers may also wish to consult the monograph *Visual literacy in communication: designing for development* by Anne Zimmer and Fred Zimmer, which appears in this series.

EVALUATING WRITTEN MATERIALS

Functional literacy workers can become involved in testing primers, graded readers and follow-up books. Developing reading material involves words. Choices must be made concerning the words to be included in the primer and those to be used in graded books and follow-up literature. Linguists may recommend one list of words and one order of presentation; the extension specialist or industrial educator working on a functional literacy project may suggest a completely different list of words and quite another order. Without pausing to explain their conflicting viewpoints, we shall simply point out that word lists may be required for different purposes, each demanding a particular approach. Project evaluators are frequently asked to develop such lists.

Words may have two levels of meaning: denotative and connotative. 'Black' denotes a particular colour, but when used in India to describe someone, it may be interpreted as a personal insult; in other parts of the world it may have racial connotations. Health extension workers teaching family planning have found that sex-related words have strong connotative meanings. Their emotional charge is so great that denotative meanings are totally obscured. Some simple tests are available for use in testing the connotative meanings of words. One simple technique is suggested in Kenneth L. Baucom's *The ABCs of literacy: lessons from linguistics*, published in this series of monographs (see list at front of book).

The communicative effectiveness of written materials can be tested, at various levels of sophistication, by asking adults to read those materials and noting their reading speed and comprehension. Literacy workers may be interested in the over-all success of a primer, reader or follow-up book. They want to know: Does the material teach? At what level does it teach? How long does it take? Does it lead to improved reading of unseen materials? Sohan Singh, in the monograph cited in the previous chapter, suggests ways of using the workbook with groups

of adult learners to provide feedback on the structure of the primer, on the level of difficulty of various individual lessons and on the reader's over-all effectiveness in teaching reading.

Measurement of changes in attitudes

Functional literacy workers all over the world, either formally or informally, ask 'attitudinal' questions such as these: Can a farmer really do anything to produce more out of the land he has tilled year after year? Should a farmer's son go to the city to work in a factory? Should a father get into debt to give his daughter a decent marriage and dowry? Should people try to limit their families? How should a government official behave towards villagers?

Functional literacy workers try, of course, to bring about attitude changes in those taking part in a programme. They constantly hope that adult learners will discard old attitudes that stand in the way of progress. They wish adults to become re-socialized, acquire new identities, and become active, self-reliant and forward-looking citizens.

TESTING FOR CHANGES IN THE SPECIFIC ATTITUDES TAUGHT

Testing for changes in attitudes is complex. We cannot see people's attitudes directly. We must either question them or judge from an analysis of their behaviour and actions what their attitudes may be. When we question people about their attitudes, they do not always understand what we wish to know; when they do understand, they do not always disclose their real attitudes. They may tell us what is socially acceptable or what we want to hear.

Also, as indicated by the taxonomy of educational objectives in the *affective* domain, changes in attitudes and values manifest themselves at various levels and to different degrees. For example, one may change one's attitude and yet lack the courage to act according to the newly-acquired attitude. Nevertheless, literacy workers must find a way of testing changes in those specific attitudes that were deliberately taught as part of the programme. They will probably find it useful to make a distinction between attitude change that was merely expressed and attitude change that was translated into action.

TESTING FOR CHANGES IN GENERAL ATTITUDES

It has often been suggested that literacy leads to changes in *general* attitudes, even if these have not been deliberately taught as part of the

course. Professor Alex Inkeles and his associates tell us that literacy contributes to the over-all modernity of the literacy class participant. They have defined the modern man as follows: he becomes an informed, participating citizen; he develops a sense of personal efficacy (that is, he begins to feel that he himself can be effective in changing things in the world around him); he becomes highly independent of traditional sources of influence; and he becomes open to new experiences and ideas. He develops a new approach to time; accepts both personal and social planning; begins to respect rules; and respects the rights of persons dependent upon him both at home and at work.

Tests are available for evaluating changes in individual modernity (see the *Dossier of instruments*). Elaborate testing of changes in general attitudes should, however, be entrusted to specially-trained evaluation specialists.

Evaluation of teacher/supervisor effectiveness

Functional literacy workers may often wish to evaluate teacher effectiveness. Since similar approaches can be adopted in evaluating teachers and supervisors, we shall confine our comments to the evaluation of teacher effectiveness.

One approach could be described as 'anticipatory evaluation'. This assumes that those teachers who make the best use of teacher-training will probably be the most effective teachers. Accordingly, testing concentrates on information retention and skills learned during teacher-training. Objections to such anticipatory evaluations of teacher effectiveness can easily be imagined.

EVALUATING TEACHER BEHAVIOUR BY OBSERVATION

Another approach to evaluating a teacher's effectiveness consists of observing and evaluating his behaviour in classes and with adult groups. If the teacher displays good teaching behaviour, he is considered to be effective.

Before using such an approach, the following points must be clarified:

1. What is teaching, what are its components, and how do these components appear as observable behaviour?
2. Which combinations of behaviour are good teaching and which are unacceptable? What are the criteria for making such judgements?
3. Is it possible to observe teaching behaviour in the class or within an

adult group with sufficient objectivity to make judgements about it? Is it possible to record it for later examination with regard to particular teaching events and teaching episodes for the same purpose?

Existing techniques for the observation and evaluation of teacher behaviour can easily be adapted for use in evaluating the effectiveness of teachers in functional literacy classes (see the section 'Observation' in Chapter Seven).

EVALUATING TEACHER EFFECTIVENESS BY TESTING LEARNER ACHIEVEMENT

Teacher effectiveness is occasionally assessed by evaluating learner achievement, assuming that if the learners have not learned, the teacher has not taught. While the weight of impact can be persuasive, educational programmes are often interested in both means and ends. Wherever possible, therefore, it is advisable to evaluate teacher behaviour in conjunction with the impact of teaching on learners.

TEACHER EVALUATION: A BROADER PERSPECTIVE

A broader approach to evaluation was adopted by the present author in an evaluation study of teacher effectiveness.[1] Literacy teachers who had taught for a few months were tested for reading comprehension, arithmetical skills and visual literacy. They were also questioned about their emerging consumption patterns, and any changes in their status and aspirations. Thus an awareness of what was happening to the teachers as persons and professional workers was developed. This profile generated important ideas about teacher effectiveness and changes in the lives of the teachers themselves.

Evaluation of systems of action

The change-makers—functional literacy workers, extension specialists and adult education organizers—seldom think of evaluating their own competence, the quality of their own decision-making processes, and

1. H. S. Bhola, *Literacy teachers of adults* (An evaluation report on training and performance of voluntary literacy teachers in a functional literacy project in the United Republic of Tanzania). Mwanza (United Republic of Tanzania), Unesco/UNDP Work-Oriented Adult Literacy Pilot Project, 1970. 101 p. Reprinted as *Occasional papers in reading*, Bloomington, Indiana, School of Education, Indiana University, July 1971. A long article based on the monograph appeared in *Literacy discussion* (Tehran), vol. 2, no. 2, Spring 1971, p. 109–48.

their own planning and organizational decisions. They fail to see ambiguities in the roles they create, conflicts in the groups within which they work or breakdowns in the organizations they develop.

Tests are available to evaluate the literacy teacher's inter-personal skills, his role-related tensions and his values with regard to this work. Tests are also available to evaluate organizational climates and management effectiveness at the project office level. It is also possible to evaluate the change models used by decision-makers and the decision-making skills of literacy specialists. For example, one can evaluate the level of bureaucracy of a functional literacy project to see whether it is likely to destroy local initiative or whether there is room for some sort of spontaneity. One can also ask questions about the literacy organization itself and the new secular institutions it has introduced into the community: Will the co-operatives or *panchayats* set up under the functional literacy programme remain in existence once the project comes to an end? Have they become institutionalized?

These are important questions. However, since they are rooted in the theory of organizations and the related concept of institution-building, discussion of them will be reserved for a separate monograph on the mobilization, planning and organization of literacy work.

Cost-benefit and cost-effectiveness analyses of literacy

Cost-benefit analysis is an accepted component of most functional literacy projects today. In this age of accountability, people want to find out what their money has bought.

Nevertheless, cost-benefit analysis is far from simple and can turn into a complex exercise in arbitrariness and self-deception. One can usually estimate the costs of a programme in terms of cash, although many economists assert that this is not easy. Planners can estimate project expenditure on things such as wages and rents, but are not always able to include the cost of all the facilities used or that of 'piggy-backing' on a multitude of officials assisting a particular programme. It is also hard to put a money value on human and social costs. How much does it cost to sit in embarrassment among a group of people you have never met before? How much does it cost to send children to school instead of keeping them at home to work or take care of younger siblings? What is the cost of changing from one cash crop to another? What is the cost of uprooting oneself from the native

village in order to work in a factory? Or of surrendering claims to a small piece of land in order to become part of a large co-operative?

Benefits are even harder to compute in money terms. One can put a cash value on increased productivity or fewer sick days, but one cannot put a cash value on the education of children, change in women's attitudes, and development of new identities among farmers, workers and housewives. Any arbitrary attempt to do so soon reveals the utter futility of cost-benefit analysis. Different analysts will end up with different ratios, based on guesses which may be politically coloured.

Because of the intangibility of cost-benefit analysis, many functional literacy workers prefer to conduct cost-effectiveness studies. Instead of asking what benefits were obtained at what costs, they try to discover whether the money was spent effectively. This presents a different set of questions that can be handled more honestly.

Community surveys for the study of impact

Did the functional literacy programme in any way affect people other than those who actually took part in it? Did an accumulation of learning by individuals in the community emerge as community learning? Did the community and its institutions change? These are the types of question that a community survey at the end of a functional literacy programme can help answer.

Here, we must repeat our earlier remarks concerning base-line surveys. Without a good initial base-line survey it is hard to draw valid conclusions from a community survey of impact at the end of the programme. Indeed, a good community survey of impact should relate directly to the base-line survey undertaken at the beginning of the project. Two further points are worth noting:

Latency periods: Instructional inputs are like seeds in the ground; they need time to sprout and grow. Depending upon the nature of the process involved (teaching, animating, role teaching, re-socializing, organizing) there may be long or short periods during which intended effects remain latent. In studying the intended impact of programmes, community surveys should take into consideration latency periods for different formative processes and social actions.

Unintended and unanticipated consequences: Successful social interventions produce the expected results, but may also have unintended

consequences. A functional literacy project may succeed in obtaining piped water for village homes, but in so doing may accidentally remove the only opportunity for women to enjoy social interaction by meeting at the village well. Another project may succeed in assisting farmers to earn more, but the money so earned may be put to unproductive uses—drinking, gambling, funeral feasts, and buying more cows or wives. Some consequences of change, while unintended, may be anticipated: a study of existing social patterns in Sukumaland in the United Republic of Tanzania should enable the literacy worker to predict that increased future earnings will be used to buy more cattle. Other consequences may, however, be totally unforeseeable. Evaluators of functional literacy projects should design evaluation instruments flexible enough to discover, and accommodate examination of both unintended and unanticipated consequences (see the section on 'Goal-free evaluation' in Chapter Three).

Summary

In this chapter, we have outlined some of the tasks that functional literacy evaluators are frequently asked to perform. Our intention has been to acquaint evaluators with these tasks, and their objectives, and to show how the methods and protocols of testing and measurement discussed in earlier chapters may be used to generate data for these special evaluative purposes.

CHAPTER TEN

Inspecting and analysing data: the use of statistics

This chapter on statistics is both elementary and brief. The main purpose of including this chapter in the monograph has been to point to some of the possible uses of statistical techniques in literacy evaluation. Thus, only the barest minimum of statistical concepts has been introduced to the reader. To follow up on the study of statistics, the reader is referred to an introductory textbook on statistics.

In the two previous chapters, we have been concerned with generating and organizing data, using various tools of measurement and mechanisms for storage. But how to make sense of these data once they are available? How to draw some confident conclusions?

We must now concern ourselves briefly with the procedures that can be used to make data speak to us. The data must tell us what is better than what else and by how much. They must reveal to us the patterns and relationships that were not visible before. Some of the analysis of available data can be done visually but some must be done statistically. In visual analysis, as the name implies, we simply look at the data and from this examination attempt to make statements about trends that seem apparent. This type of analysis is, of course, both difficult and tricky when it involves large numbers of test scores or other such data. In the latter case, statistical analysis would be necessary.[1]

What is statistics?

Statistics, simply, is the name given to the science of dealing with the collection, analysis, interpretation, and presentation of masses of numerical data. Statistics teaches us to look at numbers carefully, to know how they come about, and to put appropriate conceptual

1. Material for this chapter was originally put together by Dr. John J. Hanitchak and reviewed by David P. Rutt, Jr. The current version is a further revision of the material.

frameworks around such sets of numbers. Statistics cautions us against the snares of human perception and human logic and against the false persuasiveness of numbers. By looking at half a dozen Colombian farmers who are unable to read, we should not conclude that Colombian farmers do not read. Yet many of us do often jump to such conclusions. Many of us do not see through the bluff of statement which asserts that a gentleman is 50 per cent trustworthy!

Standardizing scores for comparisons

Literacy workers often want to compare an adult learner's previous performance in reading with his current performance in a general way. They may, that is, compare the learner's reading score from an earlier test with the new score in the current test. But is the score of 30 in the current test better than the score of 25 in the earlier test? It may be, but we cannot be sure. Many complex issues are involved here. At the very least, we will have to ask: When the score was 25 what was the possible maximum score? Again, when the score was 30, what was the possible maximum? If the adult learner in the first test scored 25 out of 50 and in the second test scored 30 out of 100, then 25 is better than 30. Before scores can be compared, they must be *standardized.* The easiest way of standardizing literacy scores would be to convert them into percentages. Simple visual comparisons can then be made between the scores made by an adult learner in two tests:

$$\text{Scoring in the first test: } \frac{25}{50} \times 100 = 50\%$$

$$\text{Scoring in the second test: } \frac{30}{100} \times 100 = 30\%$$

Clearly, the performance for the first test was relatively better than the performance in the second test.

Rank ordering for comparisons

One way of comparing adult learners in a literacy class would be to rank order them according to the scores made by each adult learner. For example, eleven adult learners in a literacy class may have scored as follows: 23, 11, 21, 19, 17, 21, 16, 16, 13, 21, and 15.

They may be ranked as follows:

Scores	*Ranks*
23	1
21	3
21	3
21	3
19	5
17	6
16	7·5
16	7·5
15	9
13	10
11	11

Where scores are tied, we assign to all those tied scores the rank that falls in the middle of the ranks of the preceding and the following scores. Thus all three scores of 21 get the rank 3, which is the mid-point between 1 and 5. Both scores of 16 get the rank of 7·5, which is the mid-point between 6 and 9.

Frequency distribution

If you had a class of forty adult learners, how would you describe this group with respect to achievement in mathematics? You could list all the scores individually but that might not tell you anything more than you knew already. A first step in organizing the data would be to list each score on the test and count the number of people who attained that score. That would give a frequency distribution. For example:

Frequency distribution of a mathematics test of 20 items

Score	Frequency
16	3
15	6
14	4
13	6
12	5
11	2
10	4
9	4
8	4
7	2
	40

This is much easier to read than a simple listing of individual scores and gives us a general idea about relative achievement within the group of adult learners. It is possible to simplify these results even further by using intervals to reduce the number of categories, as in the following table:

Frequency distribution by intervals

Score intervals	Frequency
15–16	9
13–14	10
11–12	7
9–10	8
7–8	6
	40

We can see that reducing the data in this manner makes it easier to interpret the data. However, we do lose information. If only interval data are available, we are not able to indicate how many students obtained a score of 9 or 10. We know only that eight students achieved a score of 9 or 10 points. It would be important for literacy workers in most cases to keep a note of the original scores (also called disaggregated data) so that various other statistical operations could be performed if necessary.

At the beginning of this chapter we talked briefly of visual analysis of data. Let us see how the mere organization of data for visual examination can sometimes produce useful insights.

Visual analysis—some examples

Figures 1 and 2 on the next page provide examples of one way of organizing data for visual analysis. The data have been arranged into a table in terms of various categories and have thus been, in a sense, analysed. Such cross-tabulations are often referred to as 'crossbreaks'. A cross-tabulation of various factors will allow us to interpret the relationships between these factors and will give some of the answers we might need from the data.

For example, we may look at the performance of each individual learner on each separate item of unit test. Under Lesson One, notice

Lesson	I					II					III				
Question Number	1	2	3	4	total	1	2	3	4	total	1	2	3	4	total
Total possible points	5	5	5	5	20	4	5	6	5	20	6	5	3	6	20
Learner A	5	3	2	1	11	3	2	2	1	8	4	4	3	5	16
B	5	3	4	4	16	4	0	5	4	13	5	5	3	4	17
C	4	2	5	4	15	4	4	5	5	18	6	4	3	5	18
D	5	3	1	2	11	2	3	2	1	8	4	4	3	5	16

FIGURE 1. *Data on a mathematics test*

Lesson	I			II			III		
Attendance	High	Medium	Low	High	Medium	Low	High	Medium	Low
Learner A	11				8				16
B		14				13		17	
C			15			18			18
D	11			7					16

FIGURE 2. *Attendance and final grades crossbreak*

that all learners did equally well on Item 1, and equally poorly on Item 2. On Items 3 and 4, no pattern seems to emerge. Thus we might conclude that Items 3 and 4 are poor items and should probably be reworked or replaced. Or maybe Items 1 and 2 are not good items, since they do not seem to separate good students from bad. That is, they do not discriminate.

We could also compare learners with regard to their performance within lessons or across lessons. Thus, in Lesson One learners A and D scored the lowest in the group, and across the next lesson the same

appears to be the case. But within Lesson Three a change takes place. What might be the reason for such a change? Actually, it is anybody's guess. It might be the introduction of a new teaching method or some other factor. It is only through the manipulation of such factors (by using an appropriate evaluation design) that the important factor or factors can be separated. This topic will be discussed later.

Figure 2 is another example of a crossbreak. This time, we are comparing the final grade within each lesson with attendance by the individual student. Study these data carefully and see if you can make some inferences from a visual analysis of them.

Statistical analysis

The second type of analysis we might perform is statistical. Statistical analysis, most of the time, can even tell the analyst the level of confidence with which he or she can make inferences. For example, ·01 level of confidence means that the inference has only one in a hundred chance (or probability) of being wrong. The confidence level of ·001 means that the inference has only one in a thousand chance (or probability) of being wrong.

Comparison between groups

Even when literacy workers have no desire to do any sophisticated statistical analyses, they are often faced with the need to compare one functional literacy group with another on some characteristic. A supervisor may want to compare the performance of two groups being taught by two different teaching methods. To compare them as groups, the supervisor will have to think of some *central tendency* on the basis of which the two groups could be compared. There are three statistics which provide a measure of the central tendency of groups: *mean*, *median* and *mode*.

Arithmetic mean

The arithmetic mean is calculated by adding all the scores in a distribution and dividing the resulting sum by the number of cases added. For example:

Score (x)	Frequency (f)		fx
16	2	(16×2)	32
15	7	(15×7)	105
14	4	(14×4)	56
13	6	...	78
12	6	...	72
11	2	...	22
10	4	...	40
9	4	...	36
8	4	...	32
7	1	...	7
	n = 40		Σ fx = 480

Mean ($\bar{X}$) = 480 ÷ 40 = 12·00
The arithmetic mean for our distribution is thus 12.

The *median* is another measure of central tendency. The median identifies the score that divides the group in two. Thus, 50 per cent of the obtained scores are above this score and 50 per cent are below it. In our example, we are looking for the score that will divide our forty scores with 20 above that score and 20 below. In our distribution we find our 20th student opposite the score 12. There are 6 students with scores of 12. All scores between 11·5 and 12·5 have been located in this category and it will take 5 of the 6 scores in this category to give us twenty scores. Thus, 5/6 of 1·0 (the difference between 11·5 and 12·5) is ·83 and 11·50 + ·83 = 12·33, which is the median score for this distribution of tied scores.

If the data are an odd number of untied scores such as 12, 15, 19, 24, 27, the median score is 19. On the other hand, if the data consist of an even number of untied scores such as 14, 19, 21, 24, the median is calculated by determining the midway point between the two central values (19 + 21/2 = 20).

The last measure of central tendency is the *mode*. This is the most easily obtained measure, for it is simply the number that appears most often within a distribution. In our example, the mode is 15. It should also be kept in mind that in a given distribution there can be more than one mode.

With a measure of central tendency we have a way of looking at the group in general terms; the score most typical or representative of the

group. In addition, we have a way of making approximate comparisons between individual scores and the group as a whole. Cecilia, with a score of 14, is in the top half of our example group distribution and Jean, with a score of 9, is a little below average in the lower half of the group. We have not only a way of describing the group with some efficiency, but also a way of thinking about individual scores.

The ease with which measures of central tendency are calculated can be a mixed blessing for it can lead to an inappropriate use of these statistics. The mean is affected by all scores in a distribution, especially if they are extreme. For example, if we knew five men, of whom four made $100 per year and the fifth $10,000 per year, it would be totally inaccurate to state that the average income of these five men was $2,080. The median, again, is not influenced at all by the *size* of the score added to a group, but can be influenced by a change in the *frequency* of a given score. Finally, the mode is probably useless for a measure within a small group because it is subject to unstable shifts with the addition of extreme scores.

If the distribution of your data is even and has only one mode it is probably safe to use any of the statistics mentioned. If there is any doubt, you should consult a statistics textbook.

Variability. Knowing only a measure of central tendency about a group still provides us with limited information. Can we say that two groups with the same median on the same test are equal? We cannot. We need more information. Examine the following examples:

GROUP 'A'		GROUP 'B'	
Student	*Score*	*Student*	*Score*
A	7	D	10
B	6	E	6
C	5	F	2
Total =	18	Total =	18

Group 'A': n = 3, Mean = 6
Group 'B': n = 3, Mean = 6

We have the same number of students and the same mean but we find that the two groups are quite different. The difference is apparent in the spread or *range* of the scores. Group 'A' has a range of 2 (5 to 7)

and Group 'B' has a range of 8 (2 to 10). If, for example, the test in our illustration represented a test of literacy with 10 items, we can see that the students in group 'A' are relatively the same, or about 'average'. However, Group 'B' contains students at the 'superior', 'average', and 'poor' levels. A teacher would have to work with these groups in very different ways.

Median and mode have been described as 'terminal statistics'. Once they have been computed, there is nothing more that can be done with them. The arithmetic mean, however, is a very useful statistic and can be used in many useful statistical operations. First, the mean can be put to use to determine the Standard Deviation (SD) of a distribution or set of scores. SD is an indicator of the scatter, dispersion or variance of scores in a distribution. The means of the two groups of scores may be the same. But this may hide important differences between those two groups. SD can bring out those differences in the two groups, if such differences do exist. Secondly, the mean can also be used to work out the 'standard error of the difference between means' and be put to work in applying the t-test. The t-test answers the classical question: Are two groups (one that used the special functional literacy primer and the one that did not) really different in reading performance?

It is not within the scope of the present chapter to describe and discuss these statistics or to demonstrate methods of computing them. Readers are referred to an introductory text on elementary statistics.

Yet another statistic that should be useful for the literacy worker is the Chi square. The type of question it can answer is: Are the frequencies of responses to the question (Does the community need family planning education?) obtained from two different communities really significant in allowing us to think of those two communities as having different attitudes? Here again, the interested literacy worker should consult a text of elementary statistics.

Correlation

Another type of descriptive statistic frequently used is the *coefficient of correlation.* This statistic describes the relationship between variables, and could thus assist us in answering the question: 'What is the relationship between achievement in mathematics and the ability to read?' The coefficient of correlation provides us with a single number that gives some indication of the *degree* of relationship that exists.

For example, suppose we were interested in determining the rela-

tionship between achievement in mathematics and reading for a certain group of students. The basic data we would have to work with are the results of tests in mathematics and reading for the group:

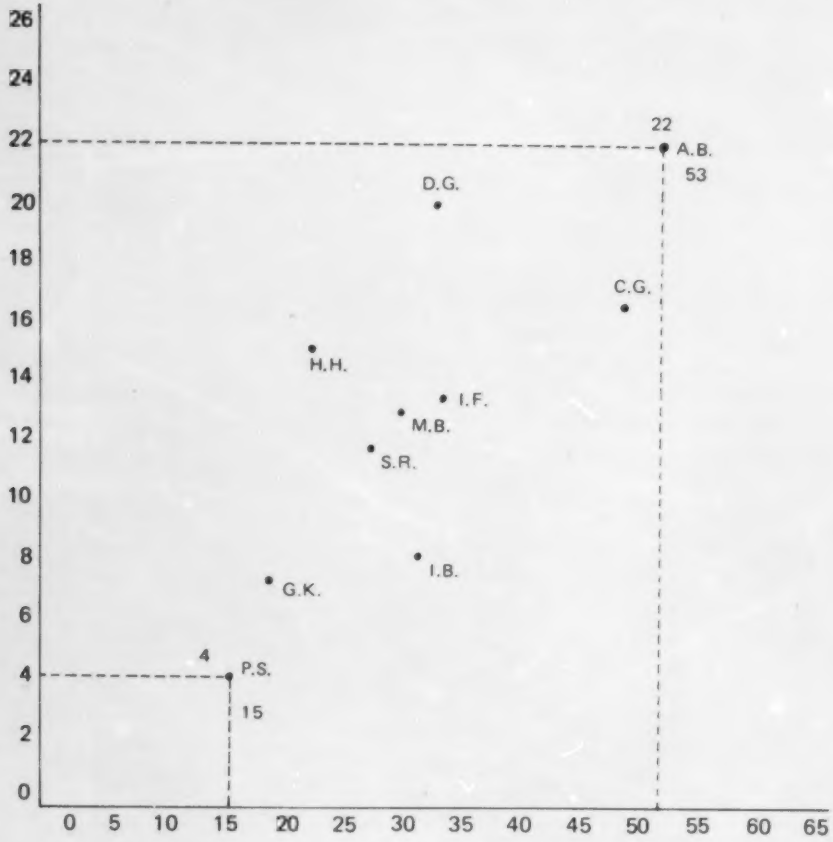
Student	*Score on maths test*	*Score on reading test*
A.B.	22	53
D.G.	20	32
C.G.	17	48
I.F.	15	36
H.H.	15	22
M.B.	14	30
S.R.	11	28
I.B.	8	32
G.K.	7	18
P.S.	4	15

Look carefully at the data. It is difficult to discern the actual relationship between reading and maths scores. There appears to be a tendency for the high scores in maths to correspond with high scores in reading, and low scores in one are also matched with low scores in the other, but we cannot give a value to that relationship from such cursory inspection. To visualize this information more clearly we could place it on a graph such as that shown on the opposite page.

Each point on this graph represents two scores for each student, one for mathematics and one for reading. For example, Student 'P.S.' received a score of 4 in maths and 15 in the reading test. His 'point' is located in the lower left-hand corner of the graph. The two grades of Student A.B. have been located in the same fashion in the upper right-hand corner of the graph. Each set of scores for the other student has been located on the graph in the same manner. Looking at the general flow of the scores we can see that they 'move' from the lower left-hand corner to the upper right-hand corner, which indicates a high positive relationship. That is, the high scores on the maths test also scored high on the reading test.

When we go about graphing data to get an idea of the relationship (positive, negative or none) we can then visually inspect the graph. There are generally three possible results that we are able to obtain, as demonstrated opposite.

In case A we have a result similar to that encountered with our reading and maths data. The scores group around the diagonal line that was drawn from the lower left-hand corner to the upper right-hand corner. High scores in maths tend to be found with high scores in


26
24
22
20
18
16
14
12
10
8
6
4
2
0
0 5 10 15 20 25 30 35 40 45 50 55 60 65
22
A.B.
53
D.G.
C.G.
H.H.
I.F.
M.B.
S.R.
I.B.
G.K.
4
P.S.
15

A
B
C

reading. Also the 'tighter' the data points cluster along the line the more perfect the relationship and consequently the higher the relationship.

In case B we find the opposite situation. The results of another reading and maths test cluster around a diagonal line that moves from the upper left to the lower right-hand corner. This indicates a negative correlation, so that a person who scored high on one exam probably scored low on the other. Again, the tighter the data points fit to the diagonal the higher the negative correlation.

In case C we see that the data points do not group themselves around either diagonal. The scores appear to be evenly dispersed, that is, a high scorer on the maths test could have scored high, medium or low on the reading test. Such a result would indicate *no* relationship between the variables.

Even though the graphing of our data provides us with a generally efficient way of determining relationships between variables, it still lacks precision. It is possible to assign a number to the relationship to give a clearer picture. Correlation coefficients are calculated so that the obtained value may range from (−1·00) through zero to (+1·00). We can encounter any value between those extremes: −0·68, −0·46, −0·05, +0·24, +0·79, +0·97. It should be pointed out that the use of a negative or positive sign indicates only the magnitude of the relationship that exists within a set of data *not* the 'goodness' or 'badness' of that data. Thus a correlation coefficient of −0·67 is a stronger relationship than +0·16. Also with regard to what a high correlation or a low correlation is, we suggest the following rule of thumb:

correlations between 0·00 and 0·30 are low
correlations between 0·30 and 0·60 are medium
correlations between 0·60 and 1·00 are high.

It should be kept in mind, however, that these guidelines are strictly arbitrary and should be used in a very general fashion. Low, consistent correlations between two factors, for example, may indicate an important relationship and therefore should not be ignored.

There are various ways to calculate a correlation coefficient depending upon our purposes and the characteristics of our data. The 'rank order correlation' or 'rho' is one useful method for calculating a correlation coefficient with small amounts of data up to 40 cases. The

formula is:

$$\rho = \frac{1-6\Sigma d^2}{n(n^2-1)}$$

where Σd^2 = sum of d^2's

d = the difference in rank

n = the number of cases

As the following example will show, only elementary school arithmetical skills are needed for calculating rank order correlations (rho). Using the data from above we will illustrate the use of this formula.

Student	Maths Score	Maths Rank	Reading Score	Reading Rank	d	d^2
A.B.	22	1	53	1	0	0
D.G.	20	2	32	4·5	2·5	6·25
C.G.	17	3	48	2	1·0	1·00
I.F.	15	4·5	36	3	1·5	2·25
H.H.	15	4·5	22	8	3·5	12·25
M.B.	14	6	30	6	0	0
S.R.	11	7	28	7	0	0
T.F.	8	8	32	4·5	3·5	12·25
G.K.	7	9	18	9	0	0
P.S.	4	10	15	10	0	0
						$\Sigma d^2 = 35·00$

$$\rho = 1 - \frac{6\,(35)}{10\,(100-1)}$$

$$= 1 - \frac{210}{990}$$

$$= 1 - 0·21$$

$$\rho = 0·79$$

In the table above you observe a column for scores and a column for ranks. What has been done is to convert the real scores to ranks, 1 for the highest score, 2 for the second, etc.

Column *d* is only the difference between the corresponding ranks of the maths score and reading score for each person: d^2 indicates that the value *d* is then multiplied by itself (squared). This value is summed for each rank, them multiplied by 6 and divided by the product of the number of people (n) multiplied by the square of n minus one. This entire value in turn is subtracted from 1.

With a rho of 0·79, we are able to indicate that there is a high positive correlation between the scores in mathematics and reading for our sample of students. Those students who tend to do well in mathematics also tend to do well in reading; and, conversely, those who do poorly in maths tend to do poorly in reading. With the use of a correlation, one very important caution is in order, *correlations do not indicate causes*. In our example above we cannot say that knowledge of mathematics *causes* high reading scores or that the opposite is true. We can only establish the magnitude of the relationship. The problem of cause is able to operate in three ways:

1. Factor A causes factor B
2. Factor B causes factor A
3. Some third factor C (or other unknown factors D, E, F, etc.) influences the relationship or performance in A and/or B.

For example, some form of intelligence (factor C) may be a factor in achievement generally and therefore influences ability to learn mathematics, reading or other subjects.

The evaluator with questions concerning students and educational situations can obtain a variety of data that will enable him to answer them. These data, accumulated systematically over a period of time—a year or several years—are probably more valid and useful in assessing directions of change and for judging educational activities. However, all information should be used with a degree of caution. It is, perhaps, preferable to use information to raise questions rather than attempt to arrive at immediate conclusions. But eventually within an evaluation effort we may want to begin to answer such questions as: 'What factors are influencing the learning of my students?' or 'Will the same results be obtained with other age groups?' When we begin to ask these types of questions we can no longer be satisfied with simply collecting data in a random, haphazard fashion; on the contrary, it must be collected in a systematic and controlled manner. If we were interested in further exploration of the maths reading issue we should have to establish a situation that would control other factors that could influence the

relationship between these two variables. This means experimentation so that we can be more confident in establishing a valid relationship between reading and maths.

Summary

In this chapter, we have included a brief introduction to some basic statistical concepts. The uses of frequency distributions and cross-breaks in organizing data to generate information are shown. Statistics that reflect the tendencies of groups of scores are explained, so that comparisons between groups can be made. Correlation, an important statistic for literacy workers, is also discussed.

CHAPTER ELEVEN

Evaluation design: increasing the validity of conclusions

The term 'design' denotes deliberate, purposeful planning—a planned intention. Design governs our functioning in life and work; it is a mental scheme in which means to an end are laid down. From this definition, it follows that all functional literacy evaluators must have an evaluation design if they are to do their job properly.

Good evaluation design will provide data that will help decision-makers to draw valid conclusions. It will help them to defend the validity of their perceptions about the present and their judgements about future programme actions. It will promote confidence among both clients and colleagues about the direction the programme may be taking and the progress made.

Table 2 lists a number of statements that could conceivably be made by a literacy worker and shows how, in all good faith, other people might cast doubts on their validity.

These are, of course, only a few examples. Many other such assertions could be similarly challenged. Three types of doubts are illustrated by the Table: (a) the adults selected for study may not be representative of all the adults in the community; (b) differences between groups ascribed to project action may have existed from the start; (c) external influences may have produced change. In each case the *validity* of conclusions has been questioned.

Professors Donald T. Campbell and Julian C. Stanley have named eight factors that may threaten the *internal validity* of experimental (evaluation) design, and four that may affect its *external validity*. Evaluators should find their list instructive.

INTERNAL VALIDITY

1. *History:* An outside influence—such as a presidential speech, a famine, the opening of a factory—may become an important variable.

Table 2. *Challenging the validity of statements*

ASSERTIONS	OBJECTIONS
Adult attitudes towards literacy have changed drastically because of the project.	Maybe they have changed not because of the project, but because of the President's speech on national radio. Maybe they have changed not because of the project, but because the newly-opened textile factory has declared its preference for literate and semi-literate labour.
The good readers in adult literacy classes later become village leaders.	Maybe this is not because literacy creates leaders but because only the social classes already in power are being attracted to literacy classes.
Adult women learn arithmetic faster than men.	Maybe the women's class had a much better teacher. Maybe all the women were employed on a semi-industrialized agricultural estate, while all the men were landless labourers.
The functional literacy programme increased the productivity of the village by 15 per cent in two years.	Maybe this was a progressive village to begin with. Maybe another village without a functional literacy programme achieved a similar increase in productivity.

2. *Maturation:* Those being interviewed or tested may change in various ways. They may become older, hungrier or wearier.
3. *Testing:* The first test may teach the items on the test and some related information. The same test (or an equivalent test) given later may not measure real changes in learning.
4. *Instrumentation:* There may be no changes in the reality being studied. Evaluation instruments may have been changed, or different observers and scorers may have given different scores to the *same* reality.

5. *Statistical regression:* This is a statistical phenomenon. If two groups score extremely high and extremely low on a first test, on a second test both will regress towards the mean. That is, those who scored extremely high will now have slightly higher scores. Such changes are quite independent of programme implementers' performance in adult classes or village communities.
6. *Selection:* Biases in selecting adults for class enrolment, testing or interviewing may make the validity of conclusions questionable.
7. *Experimental mortality:* Adults within the programme at the outset may drop out or move away. Those who remain may not truly represent the community.
8. *Selection-maturation interaction:* The peculiar chemistry of the selection of adults and their maturation may show changes not necessarily caused by the programme.

EXTERNAL VALIDITY

9. *The reactive or interactive effect of testing:* The pre-test may increase or decrease the respondent's sensitivity or responsiveness to certain programme treatments.
10. *Selection-treatment interactions:* The peculiar chemistry of selection and programme treatments may create interactions that falsify results and conclusions relating to cause and effect.
11. *Reactive effects of experimental arrangements:* Persons neither included in the programme nor given special information may still become exposed to programme information, in which case the validity of conclusions may be lessened.
12. *Multiple-treatment interference:* When the same group is frequently tested, or interviewed many times in different connections, results may become confused. The effect of an interview cannot be entirely erased from the minds of respondents and the first interview may influence the second in ways we fail to appreciate.

The purpose of evaluation design is to improve the validity of conclusions and of judgements made on the basis of those conclusions. All functional literacy evaluators must have a design, irrespective of the evaluation approach they advocate. 'Evaluation by illumination' must have a design; 'transactional evaluation' must have a design; 'participative evaluation' must have a design. The 'advocacy model'

needs a design. 'Modus operandi evaluation' is in itself a method or mode of operation and, consequently, a design.

Classical research designs can both help and hinder the literacy evaluator. Their love affair has seen many ups and downs. At first, since no distinction was drawn between evaluation and research, the classical research designs were used indiscriminately. It was then discovered that evaluation, unlike research, took place in a world of action, was less susceptible to control, and had different final objectives. Evaluators were encouraged to abandon the classical experimental design of the researcher in favour of quasi-experimental designs. At present it is being suggested that evaluators should use neither experimental nor quasi-experimental designs, but invent their own evaluation designs.

Unfortunately, evaluation design is still in its infancy. The design element in most of the evaluation approaches mentioned earlier is vague and ill-defined. To educate ourselves in the concept of design and to become proficient in inventing specific evaluation designs, we must emulate the research designers who have looked out beyond the laboratory to the field. Let us first examine some aspects of sampling, matching and control groups and then go on to consider 'quasi-experimental designs'.

Two important points must be stressed:

(i) Evaluators of literacy programmes should not make decisions about evaluation design without first knowing the particular aspects of the programme to be evaluated and taking into account the ideological leanings of the developmental élite. Indeed, the choice of evaluation design or designs is an important part of the process of formulating evaluation policy;

(ii) Evaluation designs should be developed in collaboration with the programme implementers and also, if possible, in co-operation with the clients of the programmes to be evaluated. This would not only result in considerable mutual learning, but also increase the 'psychological validity' of the design, the data collection instruments and the conclusions based on the data thus generated.

Sampling

Sampling is one aspect of design that warrants a brief description at this point. Functional literacy workers in large projects often have to deal with considerable numbers of adult men and women. In order to

find out how adult attitudes might be changing and what uses adults might be making of newly-acquired literacy skills or agricultural knowledge, literacy evaluators need not necessarily question *all* the adults concerned. They may obtain the information they require by *sampling* the adult men and women in the programme and making inferences about the whole from the part.

Sampling is the process of selecting a small number of units (adults, teachers, groups, fields, villages) in such a manner that they can be used to make estimates about the total number of units with which an evaluator is concerned. Two considerations are of paramount importance when selecting a sample: size and representativeness.

The evaluator is frequently concerned with 'population'. The everyday meaning of that term covers all the people—men, women and children, young and old, farmers and labourers, teachers and lawyers—living in a particular community or nation. In the vocabulary of research and evaluation, 'population' is that total group of people in whom the evaluator is interested. It may comprise, in one instance, all farmers between the ages of 25 and 45; in another, all women in the child-bearing age group; and in yet another, all the textile workers in a particular region of the country. Samples are, therefore, selected in order to represent particular populations.

The size and representativeness of a sample are interconnected. The most representative sample of a population is, of course, the population itself. The larger the sample, the more representative it is likely to be, but for a largely homogenous population, taking large samples can waste both time and money and a well-chosen sample of 2,000 may sometimes adequately represent a population of 100 million. Nevertheless, evaluators have to remember that samples must also have psychological validity; a sample may be statistically defensible but to clients and decision-makers it may seem too small, and thus unrepresentative.

Statisticians have worked out formulae whereby they can test the representativeness of a sample. They can also calculate the probable error, so that inferences can be made at particular levels of probability. If census data on the population being studied are available, the evaluator of a functional literacy project can use them to test the representativeness of his sample by repeating some of the questions previously asked in the census. Sample and census responses can then be compared and, if they are similar, the sample is acceptable. Another method of control is to draw several samples of approxi-

mately the same size from a single population and compare the information they produce. For greater economy, the chosen sample can be divided into two parts and the results for each of the two parts compared.

RANDOM SAMPLING: 'Random' means haphazard, without scheme. In the vocabulary of research and evaluation, random selection means that each item or person in the population being studied has had an equal chance of being included in the sample. Random samples, when of an appropriate size, are most likely to represent all the different characteristics of the total population. One method of random sampling is to arrange the population in some way, assign a number to each person, and then draw some numbers at random. When the total number of units involved is very large, tables of random numbers may be used.

LIST SAMPLING: List sampling is a modified random sampling method by which the population is arranged according to some rule, e.g., alphabetically, geographically or by mailing address—and then every *n*th number is selected; for example, every second or every twentieth number might be picked, depending upon the size of the population and the size of the sample being selected. The starting point in the selection process can be randomly selected to meet the criterion of equal chance of selection for each unit.

AREA SAMPLING: Area sampling makes a random selection not among individuals but among geographical areas. All appropriate units in the selected areas are then studied.

STRATIFIED SAMPLING: Sometimes the population of interest to the evaluator may have distinct stratifications. It may comprise urban and rural populations or different age groups—young, middle-aged and very old. There may also be clear-cut income differentials. In such a case, stratified sampling may be used. In accordance with proportions in the total population, samples should be randomly and proportionately drawn from each of its strata.

Matching and control groups

Matching of individuals and other units, and control groups are two other important elements of design. These will be discussed more fully

in the following section, but a brief explanation may be useful here.

Let us take one of the hypothetical assertions listed in Table 2: 'Adult women learn arithmetic faster than men', and the ensuing objection: 'Maybe the women's class had a much better teacher.' The real issue here relates to matching. The evaluator is being asked why women may have learned faster than men. Was it because of the women themselves, or was it for some other reason? It is possible that two teachers (one teaching a male group) were not matched on important characteristics. If two instructional methods were being compared, not only the teachers but also the adults in the two groups should have been matched.

Matching is, in fact, a means of control. If one of the two groups receives no teaching or extension, it becomes the control group for the one receiving teaching or extension.

Let us recall another hypothetical assertion: 'The functional literacy programme increased the productivity of the village by 15 per cent in two years.' A possible objection to this statement was: 'Maybe another village without a functional literacy programme achieved a similar increase in productivity.' The question here is one of control. By choosing a control village (matching most of its characteristics with those of the project village), this objection could have been refuted.

Some ideas on design[1]

Let us examine a number of possible designs, sometimes labelled 'quasi-experimental', that could be used within an adult literacy programme. From our previous discussion you will realize that a 'true' experiment necessitates the manipulation of an independent variable and randomness. In the quasi-experimental situation we usually have neither of these. We can, however, control *when* the data will be collected and *from whom* and also, in some cases, the random assignment of treatments. For this reason, quasi-experimental designs are useful in naturalistic settings.

DESIGN I—LONGITUDINAL TIME DESIGN: This design may be represented as follows:

$$Y_1 \; Y_2 \; Y_3 \; Y_4 \; X \; Y_5 \; Y_6 \; Y_7 \; Y_8$$

1. This material was developed by David P. Rutt, Jr., and is based on the ideas on quasi-experimental design put forward by Professors Donald T. Campbell and Julian C. Stanley.

The 'Y's stand for observations or tests that were administered, while 'X' stands for the introduction of some types of experimental variable.

This design, which is really an extension of the 'before and after' design, is probably one of the simplest and most effective available to evaluators. It can be used effectively where tests or records are kept on a regular basis, so that students are not aware that any special data collecting is taking place. This design can be used within a single classroom or across several but, if multiple classrooms are used, the students within the classrooms should be matched as closely as possible. For example, it would be pointless to compare a class containing a mixed (male-female) group of rural dwellers with a class composed of 17 year-old city women.

Many of the factors that might influence internal validity are controlled by the design itself. Since this is a time series, any effects of *maturation* could probably be ruled out because they would occur throughout the entire process. That is, we may notice a change between $Y_4 - Y_5$ with our experimental manipulation (X) intervening, but we have no reason to believe that this is due to maturation unless a similar change can also be observed between $Y_2 - Y_3$.

Testing and *instrumentation* are also controlled provided we use similar instruments, testing procedures and testing personnel. Thus, if we had taken measurements only at $Y_4 - Y_5$, we could not state categorically that our tests were not 'sensitizing' our students, but since tests were given over a period of time, any change that occurred must be due to some factor other than testing. The same goes for instrumentation.

A *regression* effect is also highly unlikely. If it were to influence our results we would see the greatest value at Y_1, with a steady decrease toward Y_8, rather than a jump between $Y_4 - Y_5$. However, in a study involving extremes, evaluating, for example, the effects of a special reading programme upon either very poor readers or very good readers, any changes occurring *could* quite possibly be due to the regression effect.

The factor of *selection* is controlled, provided that we do not attempt to generalize and that the same people are involved at each of the data collection points. If people are entering and leaving the experiment, *mortality* could influence the results. This effect would be compounded in the case of some type of group score (mean, median, etc.) or if no accurate records of absenteeism, drop-out and replacements have been kept.

Probably the major difficulty with this design is a historical one. In a 'true' experiment, we would be able to isolate our subjects from the outside world and thus manipulate their environment in respect of the variable(s) we wished to test. In a real-life situation this is simply not possible. The best we can do is to note the various environmental factors present and balance them against the results of our study as alternative plausible explanations. For example, if one were attempting to evaluate the effect of a new type of reading instruction on students' reading ability and the local library simultaneously launched a week-long campaign to encourage people to read, using advertisements, banners and door-to-door canvassing, it would be hard to ignore this as a possible rival explanation for any changes in reading scores. Other historical factors that might influence results are: local festivals; planting or harvesting activities; political rallies or meetings; pay cycles or work cycles.

The use of a longitudinal time design of the type proposed above entails various additional problems. Time is an important variable in this design. If we delay assessment of the impact of X at point Y_5, it is possible that some other event may occur; however, if we do not wait long enough, X may not have started to take effect. How long should one wait? This can be determined only by the evaluator's common sense and experience. It should be clearly established beforehand when data is to be collected and at what point the independent variable is to be introduced, in order to avoid any chance occurrences influencing results.

Another problem with this design arises from what we might call the 'reactive effect'. It has been found that when people are exposed to something new and pleasant they will react in a positive manner. If students are aware that they are being observed or treated differently from usual, their response pattern may be affected. For example, if we wished to assess some new mathematics materials, any resulting effect might stem from the fact that the materials are new and not that they are any better educationally. Thus, if this design is to be used, an attempt should be made to collect data unobtrusively by means of existing tests and school records. The independent variable introduced should not appear to be out of the ordinary and should be introduced without fuss or 'fanfare'. In many cases, it is best to give the classroom instructor as little information as possible. Otherwise, if he knows too much about your study, he may 'tip off' the learners by some type of unusual behaviour, thereby distorting the results.

Our method of controlling any of these effects would be to include a *control group,* which leads us to our next design proposal.

DESIGN II—THE MULTIPLE TIME-SERIES DESIGN:

$$\frac{Y_1\ Y_2\ Y_3\ Y_4\ X\ Y_5\ Y_6\ Y_7\ Y_8}{Y_1\ Y_2\ Y_3\ Y_4 \quad\ Y_5\ Y_6\ Y_7\ Y_8}$$

(experimental group)
(control group)

As you can see, this design differs from the longitudinal time design only by the addition of a control group, so that the experimental treatment can be assigned to either of the two groups at random. If random selection of groups proves impossible, at least some degree of equivalency must be sought.

This design is probably the better of the two. It allows us to control all the factors that might affect internal validity and also provides us with a control group. The design increases the accuracy of our interpretation by allowing for a multiple check, using a similar control group, on history, maturation, etc., before applying the special treatment within the experimental group.

Neither of the designs we have proposed offers any assurance of external validity. Without random selection, we are limited to the population that allows us to generalize our results. Actually, we are in a position to make proper inferences only about the group that has been involved in our experiment. If we hope to carry over inferences from our study to other groups, the same study will have to be carried out using these groups. It is only through this slow but thorough process that we can be sure the inferences we make are correct generalizations and not applicable only to characteristics found within a small group.

Earlier, we proposed two methods of analysis: visual analysis, using a crossbreak arrangement of the data, and statistical analysis, employing descriptive techniques. Both methods lend themselves quite successfully to the interpretation of data resulting from the use of either of the designs outlined above. It is also possible to employ descriptive statistical information within a crossbreak if there is an excess of data.

Evaluation is concerned with making decisions affecting various programmes. The decisions made must be as accurate as possible, because mistakes cost money and the people we are working for and with (our learners) deserve the best decisions we can make. The quality of our decisions depends upon the quality (validity) of our data. The combination of experimental design and statistical analysis

will provide us with the valid information needed to make accurate inferences culminating in correct decisions.

Summary

Conclusions reached by evaluators can be challenged by decision-makers, competitors and client groups. The factors that distort the validity of evaluation conclusions have been listed. The role of design in evaluation has been shown to be that of improving the validity of conclusions. The contributions of sampling, matching and control as elements of design have been discussed. Some quasi-experimental designs of interest to evaluators have also been included.

CHAPTER TWELVE

Interpreting evaluation results for decision-making

We have succeeded in creating data by means of appropriate evaluation designs, measurement tools and statistical treatment. We now have a set of conclusions in which we can place considerable confidence. What do we do next?

We may have information enabling us to answer the following questions: Is a particular primer effective or not? How motivated are adult learners? Do voluntary teachers teach better than those employed full time? Is the duration of classes too long or too short? Could the present level of programme services be achieved at less cost? Our answers must now be used as a basis for decision-making.

The decision-making process

Some hard facts about decision-making must be faced. First, decision-making is not merely a *technical* process. Although the results of an evaluation study may demonstrate the need to move in a new direction, or to modify the means of moving in that direction, the decision to do this will not necessarily follow, because decision-making is also a *political* process. Political considerations may demand that the advice of the evaluators be ignored. While new proposals may make sense to idealists, they may annoy certain interest groups, which may cease to co-operate or actively resist the adoption of new goals. Politicians may feel that the future situation could be worse than the present one and prefer to 'hasten slowly'.

Secondly, the decision-makers responsible for high-level policy decisions require much more information than an evaluation study can normally produce. Deciding whether a programme should be continued or discontinued—and, if discontinued, what alternative should be offered—can never be a purely technical affair. It has to be a political one.

If programme decisions are inevitably political, it is legitimate to ask whether evaluation is necessary. The following points may suggest an answer:

1. We have stated that decision-making is not solely a technical matter. On the other hand, we have *not* asserted that technical considerations play no part in decision-making; they do.
2. We have stated that decision-making at the higher levels of policy and planning is predominantly political. We have *not*, however, excluded technical considerations. Evaluation and research do inform political discussion. The technical point of view does enter the political argument.
3. There are scores of evaluation-related issues that are technical rather than political. The information produced by evaluators will be most useful in connection with such issues; for example, improving attendance, designing curricula, improving instruction, selecting learners, training teachers, producing materials and maintaining motivations.

What level of success is acceptable?

The question has been asked already and can be asked again: What level of success is acceptable? Which evaluation results should make one feel optimistic and which should make one feel despondent?

Learning, we are told, is a door that opens from the inside. Consequently, teaching—as the process of making learners learn—is a difficult task. Functional literacy, which seeks to teach the adult new social, economic and political skills as he learns to read and write, is still more complex and proportionately more difficult.

Motivations to learn literacy can be seen to have two components; one is psychological and the other structural. Most individuals, most of the time, are curious; they enjoy achieving and wish to succeed at what they do. If literacy programmes are technically and professionally sound, they are likely to teach effectively and to provide learners with psychological satisfactions. The structural component of motivation to learn is, however, the more important one. The adults in literacy programmes must be offered full participation in the economic, social and political structures surrounding them. They must be convinced that doors leading to improved economic rewards, to higher status, and to participation, decision, questioning and influence can be opened; and that literacy is the key that will open them.

Unfortunately, functional literacy workers do not have much control over the economic, social and political structures in the societies they serve. They cannot always offer adult learners full participation in the system. They may offer only partial affiliations or even, sometimes, nothing more than some psychological reward.

This being so, functional literacy workers seldom experience spectacular successes. To begin with, they have difficulty in motivating learners. Not many people want the services they offer. Even after a group comes into being, as many as one-third may drop out during the first two weeks. Of those who remain, only one-third may learn to read; the others may never succeed in cracking the code. The situation in literacy classes reproduces that in the formal school system, where failure and repetition of grades is staggering.

While the preceding comments should put the concept of successful literacy work into perspective, they should not be used as an escape hatch. Literacy workers must not give up. They must develop realism, not cynicism. They must not—blinded by their tears—lose both direction and commitment.

Evaluation of evaluation

One may also ask whether the evaluation of a literacy programme has been worth all the money and effort put into it. Answers to this question are by no means easy to find. Evaluation of evaluation is far more complex than the initial evaluation of projects and programmes.

This complexity arises from various sources. First, we cannot always say what shape the original programme would have taken if no evaluation had been envisaged. We have repeatedly stressed that the process of developing evaluation policy and plans entails a critical and analytical examination of programme objectives, underlying causal assumptions, and the logic of action strategies. Evaluation may thus make 'pre-natal' contributions to the programme before it is actually put into effect. It is possible that the whole direction of the programme changed when evaluation was proposed, or that serious errors and wastage were avoided. Secondly, it may not be possible to attach absolute values to the costs and benefits of all programme decisions to which evaluation may have contributed. Cost-benefit analyses may become arbitrary and artificial.

A middle-level literacy worker can, however, evaluate evaluation on the basis of, at least, the three criteria of *efficiency*, *completeness* and

focus. If all or most of the data generated by programme evaluation are used in programme planning and decision-making, the evaluation effort has been efficient; in other words, no unnecessary data were collected and no resources were squandered. Evaluation may be considered complete if all or most decisions made within a literacy programme or project arose, in part, from the available evaluation data. Evaluation may be said to have focus when at least the most important decisions made have used the data generated by programme evaluation.

These criteria cannot always be employed, but they can prevent evaluation from acquiring a life of its own, totally unrelated to programme decisions.

Applications of evaluation within the formal system

This monograph has dealt primarily with the evaluation of functional literacy programmes, with occasional references to the evaluation of programmes of life-long education and socio-economic development. What about educational evaluation within the formal education system? Does this monograph have implications for evaluating the effectiveness and impact of formal schooling?

The concepts and definitions to be found here are certainly equally applicable to formal education programmes. What is evaluation? What value questions are involved in evaluation and in evaluation approaches and methodologies? Should the pupils in the nation's schools be taught to compete—as in capitalist systems—or to co-operate—as in socialist societies? The choice made at this stage on the road to progress will determine the choice of approach to evaluation. What should the evaluation policy be? Here, again, one can evaluate literally hundreds of different aspects of the learning process and its relationship with psychological and social variables of students and teachers. What would be a reasonable evaluation agenda for the school system?

The model proposed earlier for evaluating functional literacy programmes is equally suitable for use within formal education if we start by describing the school as an 'instructional system' so that we can develop an evaluation system and an appropriate data system. The design concepts to be used to make valid conclusions are, similarly, transferable to the in-school situation.

The concepts of measurement and the testing techniques described above also apply to formal education. Needless to say, the choice of

techniques is not always value-free. One should remember that a humanist might well suggest the use of the illumination model of evaluation. The advocacy model could be used to confront teacher evaluation with parent evaluation, or even student evaluation, of teaching. Choices of approach will thus be all-important. In the area of testing, standardized tests, based on national norms, and criterion-referenced tests would obviously arise from two very different value systems and world views.

One need not, however, consult this monograph only to see how far it can illuminate evaluation problems within the school setting. Why not view the school as 'a broad-impact project'? Why not view the school as a system interacting with the social system around it (family influences on children; influences of the children of families; relationships between socialization at home and socialization at school; interactions of school learning with the economic, occupational and status systems in the community)? This monograph would then undoubtedly cover the evaluation of functional literacy acquired within the formal school systems. It would be central to the evaluative concerns of education within the formal setting.

Summary

The process of using evaluative information in making decisions is shown to be essentially political. Evaluation need not, but often can, influence decision-making, since politicians must deal with the information that becomes available. Determining norms of programme success or failure was discussed, as also was the question of evaluating evaluation. Finally, the possible applications of the evaluation strategy discussed in this monograph to the evaluation of formal education systems were pointed out.

CHAPTER THIRTEEN

Concluding remarks

In the previous pages we have discussed a number of ideas and procedures that should help you in your evaluation endeavours. We have stressed the importance of understanding the programme system to be evaluated. Since we cannot realistically evaluate *everything* within a programme, it is essential to lay down an evaluation policy. This evaluation policy will bridge the gap between the programme and the evaluation systems, and enable us to structure the latter efficiently in relation to the priorities of the former. From the evaluation system we shall in turn derive our data system, which will enable us to collect and store the necessary data for later use in the decision-making processes needed for programme implementation.

The generation of data for evaluation purposes is an important activity. We have explained the importance of developing indicators, various techniques of test-making and the fundamentals of measurement. We have also examined the use of evaluation designs and statistics as approaches to the generation of valid information.

The problem now facing us is similar to that confronting someone who has just spread out the pieces of a jig-saw puzzle. We have scrutinized each of the pieces in isolation, but before we can see the whole picture we must fit them all together.

In this final chapter, we shall try to fit the pieces together. We shall outline the steps you might take if you had just been appointed head of a team set up to evaluate the work of a literacy project in a developing region in a Third World country. Here are our suggestions:

1. Determine the purpose of the programme to be evaluated. This may appear to be a simple matter. You may be told something like this: 'The purpose of our programme is to teach people to read and write.' However, do not allow yourself to be satisfied with this response. On the surface it seems reasonable, but there are always

other purposes in a literacy programme that should be brought to light. To 'unearth' these hidden purposes may require patience, tact, diplomacy and, as we have stressed many times before, not only a good dose of common sense, but a desire to listen rather than talk. The rewards, in terms of a more effective evaluation of greater usefulness to the client, more than justify the time invested at this stage.

2. Determine the learning objectives and goals of the programme. In Step 1 we defined its purposes. Now we are re-stating these purposes in terms of the learning of our students, or changes that may occur in institutions or at community level. These first two steps, along with the sixth, will help you and your client to obtain a better understanding of the programme system as described at the beginning of this monograph.
3. Determine the evaluation policy. Essentially, this involves assessing if and what evaluation can actually be carried out. It also allows for a more accurate and effective evaluation system.
4. State evaluation aims in terms of purposes and objectives. This is a preliminary step in the development of the evaluation system. It involves determining intended outputs and attempting to identify unintended outcomes. As we have mentioned before, it is not possible to identify *all* the outcomes of a given programme system. Nevertheless, an evaluation policy can help us to identify outcomes that are of major significance.
5. Determine the indicators related to programme objectives and goals. As part of the evaluation system, we shall have to identify the impact of our programme on certain units within the context of that programme. In turn, within each unit (individual, group, institution or sub-culture), it will be necessary to identify an indicator, or a set of indicators, that will demonstrate the desired impact.
6. Select the relevant inputs, processes and contexts of importance to the programme system. The selection of these three features is just as important as the determination of outputs.
7. Select and develop data collection techniques and instruments. In establishing a data system you will need to consider how data are to be collected and stored, what data should be collected according to the selected indicators, and when it may also prove necessary to generate data by using some type of evaluation design. Finally, you may also need to test prototypes of any tests,

questionnaires or interview schedules used.

8. Determine needs for evaluation in terms of money, personnel and equipment. This is probably one of the most important steps, simply because most programmes have very little money available for evaluation. Careful planning and assessment are vital if needless waste is to be avoided.
9. Assign tasks and responsibilities. The number of people available —like that of other resources—will probably be strictly limited. Thoughtful planning is thus needed *before* evaluation begins, so that all members of your team know what is expected of them.
10. Train staff. You may find, after developing the evaluation system and deciding on the various tasks and responsibilities of your staff, that the members of your team are inadequately prepared. You have two options: to select new team members, or to train the staff you already have. It may be cheaper in terms of money, time and effort to search for new staff. You must know the abilities of the people assisting you with the evaluation. This also has a bearing on Step 8.
11. Carry out the evaluation.
12. Collect data. The data should be collected and stored as indicated in Chapter Eight.
13. Analyse data. The data can now be tabulated or subjected to statistical treatment, as discussed in Chapter Ten.
14. Interpret data. This may be one of the most difficult steps because it involves attempting to make sense out of the facts, figures and observations which you and your team have gathered throughout the evaluation period.
15. Compare final interpretations with original goals and objectives. Does the programme appear to be meeting those goals? If so, to what degree; if not, why not? No simple answers will emerge from an evaluation of any programme. It is possible that in many cases your results will simply ask more questions—without providing any answers!
16. Write the final report. Your client will no doubt wish to have a final record of the evaluation with a written statement of your interpretations, conclusions and possible recommendations. It is advisable to include in your report such things as: (a) statement(s) of the programme goals and objectives; (b) a summary of the procedures used throughout the evaluation (points of data collection, type of data collected, how data were collected, etc.); (c) a

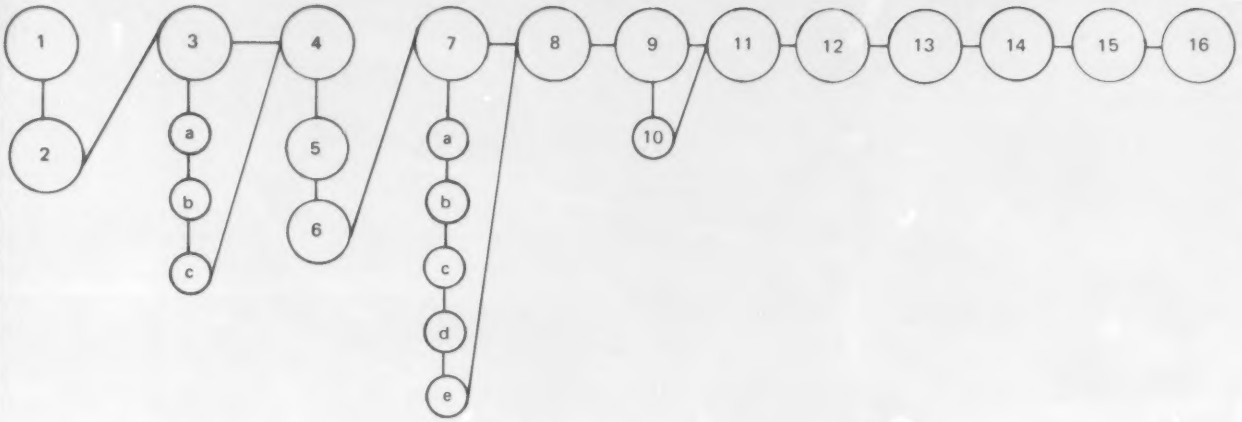


FIGURE 5. *A PERT chart for literacy evaluation.*

Notes:

1. Determine purpose of the programme.
2. Determine learning objectives and goals.
3. Determine evaluation policy:
   (a) should evaluation be done?
   (b) to what extent?
   (c) how should it be handled?
4. State evaluation objectives.
5. Determine indicators.
6. Select relevant inputs, processes and contexts.
7. Select and develop data collection techniques and instruments:
   (a) logs, tables, etc;
   (b) tests;
   (c) questionnaires and interview schedules;
   (d) evaluation design;
   (e) test prototypes.
8. Determine budgetary and personnel needs for evaluation.
9. Assign tasks and responsibilities.
10. Train staff.
11. Carry out evaluation.
12. Collect data.
13. Analyse data.
14. Interpret data.
15. Compare final interpretations with original goals and objectives.
16. Write final report.

list of the members of the evaluation team; (d) summary table(s) of the data; (e) samples of questionnaires or interview schedules used; and (f) clearly-stated reasons indicating why various decisions were taken. It is unnecessary for such a report to include everything that occurred during the evaluation. Keep the length of the report within reasonable limits, but retain all of the data originally collected for possible future analysis.

Figure 5 (p. 155) will provide you with a programme evaluation and review technique (PERT) chart for planning evaluation. When organizing your programme, you may find it useful as a reviewing device. Each of the steps discussed above has been numbered, and the sub-steps are also identified. The various steps have been arranged horizontally, from left to right, if they follow one another sequentially; and vertically if they are related or could occur simultaneously.

Summary

The material included in this training monograph has been summarized in a practical way. In this final chapter, we have looked at the problems of evaluation from the perspective of a literacy worker who joins a project and is asked to conduct evaluation. What steps might he take? In what sequence? A PERT chart has been included to help the literacy evaluator in his task.

APPENDIX I

A selected annotated bibliography

H. S. Bhola, *Social change and education.* Published by General Learning Press, 250 James Street, Morristown, New Jersey 07960. 30 pages. Price $2·50. 1976.
This instructional module clarifies the relationship between education and social change; it directs attention to the role of the élite in societies by proposing a model of cultural action as social 'affiliation/exclusion' directed by the élite.

H. S. Bhola, *Diffusion of educational innovation.* Published by General Learning Press, 250 James Street, Morristown, New Jersey 07960. 30 pages. Price $2·50. 1977.
This instructional module explains the processes of educational change; it proposes the configurational (CLER) model as a comprehensive approach to planned educational change and innovation diffusion.

Scarvia B. Anderson, Samuel Ball, Richard T. Murphy and Associates, *Encyclopedia of educational evaluation.* Published by Jossey-Bass, Inc., 615 Montgomery Street, San Francisco, California 94111. 515 pages. Price $17·50. 1975.
Includes short articles on the main concepts and techniques of educational evaluation. Each article is followed by a bibliography. Articles are arranged alphabetically.

Blaine R. Worthen and James R. Sanders, *Educational evaluation: theory and practice.* Published by Charles A. Jones Publishing Company, Worthington, Ohio 43085. 372 pages. Price $9·50. 1973.
An excellent introduction to some of the models of educational evaluation. Includes a discussion of the considerations entering the planning of evaluation studies.

PDK National Study Committee on Evaluation, *Educational evalua-*

tion and decision making. Published by F. E. Peacock Publishers Inc., Itasca, Illinois. Also available from Phi Delta Kappa, Inc., Eighth and Union, Bloomington, Indiana 47401. 368 pages. Price $10·50. 1971.
Discusses evaluation in the context of decision-making and proposes the CIPP (context-input-process-product) model of evaluation for planning evaluation studies of educational programmes and projects.

W. James Popham (Editor), *Evaluation in education: current applications.* Published by McCutchan Publishing Corporation, 2526 Grove Street, Berkeley, California 94704. 585 pages. Price $11·00. 1974.
Brings together contributions from nine American evaluation experts on various important aspects of educational evaluation.

Malcolm Provus, *Discrepancy evaluation for educational program improvement and assessment.* Published by McCutchan Publishing Corporation, 2526 Grove Street, Berkeley, California 94704. 380 pages. Price $12·50. 1971.
Includes a discussion of the development of the discrepancy evaluation model and of how the model could be used to plan and implement educational evaluation.

Robert M. Rippey (Editor), *Studies in transactional evaluation.* Published by McCutchan Publishing Corporation, 2526 Grove Street, Berkeley, California 94704. 327 pages. Price $12·00. 1973.
Advocates an emphasis on the evaluation of the processes of change and a concern with the social forces acting within the changing institutions. Case studies of what is named 'transactional evaluation' are included and an approach to a theory of transactional evaluation is developed.

Fred N. Kerlinger, *Foundations of behavioral research* (2nd edition). Published by Holt, Rinehart and Winston, Inc., 383 Madison Avenue, New York, N.Y. 10017. 7[illegible] pages. Price $14·50. 1973.
A basic text on research design, measurement and statistical analysis of data that could serve as a standard technical reference for educational evaluators.

Donald T. Campbell and Julian C. Stanley, *Experimental and quasi-experimental designs for research.* Published by Rand McNally College Publishing Company, Box 7600, Chicago, Illinois 60680. 84 pages. Price $5·95. 1966.
Classical presentation of the problems involved in the design of

educational and social experiments that sheds light on the problems involved in designing evaluation studies. Also includes a discussion of the concepts of internal and external validity.

Morris James Slonim, *Sampling.* Published by Simon and Schuster, Rockefeller Center, 630 Fifth Avenue, New York, N.Y. 10020. 144 pages. Price $2·95. 1960.
An introductory discussion, in layman's language, of sampling and related statistical concepts.

Carol H. Weiss, *Evaluating action programs: readings in social action and education.* Published by Allyn and Bacon, Inc., 470 Atlantic Avenue, Boston, Mass. 02210. 365 pages. Price $7·95. 1972.
A book of readings on the theory, politics and methodology of the evaluation of educational and social programmes.

Earl R. Babbie, *Survey research methods.* Published by Wadsworth Publishing Company, Inc., Belmont, California 94002. 384 pages. Price $8·55. 1973.
An introduction to survey research as a social science method, survey research design and analysis of survey data.

Benjamin S. Bloom (et al.), *Taxonomy of educational objectives—Handbook I: cognitive domain.* Published by David McKay Company, Inc., 750 Third Avenue, New York, N.Y. 10017. 207 pages. Price $3·95. 1956, 1975.
Distinguishes cognitive learning from affective and motor learning and presents a detailed classification of possible educational goals and objectives within the cognitive domain.

David R. Krathwohl, Benjamin S. Bloom and Bertram B. Masia, *Taxonomy of educational objectives—Handbook II: affective domain.* Published by David McKay Company, Inc., 750 Third Avenue, New York, N.Y. 10017. 196 pages. Price $3·95. 1964, 1974.
Distinguishes affective learning from cognitive and motor learning and presents a detailed classification of possible educational goals and objectives within the affective domain.

Jon Clark Marshall and Loyde Wesley Hales, *Essentials of testing.* Published by Addison-Wesley Publishing Company, Reading, Mass. 01867. 162 pages. Price $4·50. 1972.
An introduction to the principles of measurement, techniques for constructing tests and analysis of test results.

Norman E. Gronlund, *Preparing criterion-referenced tests for classroom instruction.* Published by the Macmillan Company, 866 Third Avenue, New York, N.Y. 10022. 55 pages. Price $2·50. 1973.
A brief introduction to the concept, design and interpretation of criterion-referenced tests of achievement.

Eugene J. Webb and others, *Unobtrusive measures: nonreactive research in the social sciences.* Published by Rand McNally College Publishing Company, Box 7600, Chicago, Illinois 60680. 225 pages. Price $5·95. 1966.
An excellent monograph on approaches to observation and measurement particularly suited to holistic and naturalistic methodologies of evaluation.

Ole R. Holsti, *Content analysis for the social sciences and humanities.* Published by Addison-Wesley Publishing Company, Reading, Mass. 01867. 235 pages. Price $3·95. 1969.
An introduction and guide to content analysis as an approach to documentary research and evaluation.

APPENDIX II

Glossary of terms

ADVOCACY MODEL: An approach to evaluation in which two opposing views are presented—as by advocates in a court of law.

AFFECTIVE: Relating to emotions and feelings; hence affective behaviour and affective learning. Learning of attitudes and values are also considered aspects of affective learning.

BASE-LINE SURVEY: An initial survey that can serve as a base for comparing changes observed subsequently.

BEHAVIOURAL OBJECTIVE: A statement which describes in observable and testable terms that which will be demonstrated by the student as a given learning outcome.

CONNOTATIVE: Relating to an additional meaning of something, beyond primary meaning. (See also denotative meaning.)

CONTEXT EVALUATION: Assessing and evaluating the environmental variables of a programme.

CORRELATION: An index of relationship between two characteristics (often expressed as two groups of numbers) that ranges from −1·00 through zero to +1·00.

COST-BENEFIT ANALYSIS: An assessment of the inputs and outputs of a programme in terms of their monetary values.

COST-EFFECTIVENESS ANALYSIS: An assessment of the inputs, processes and outputs of a programme in terms of the effectiveness of means employed for the ends obtained.

DATA SYSTEM: An arrangement of methods and procedures by which data resulting from the implementation or evaluation of a programme are organized and stored for use.

DENOTATIVE: Denoting or indicating the primary meaning of something.

DIVISION OF LABOUR: The breakdown of work into its component parts and its division among different individuals with a view to maximum productivity.

EVALUATION: The process of determining the worth, significance, value or quality of an item, event and/or person, usually by careful appraisal and study. Essentially, evaluation is the process of describing experience objectively to improve performance.

EVALUATION SYSTEM: An arrangement of methods, procedures and plans of action designed to provide decision-makers with information on the inputs, outputs, context and process of a given programme.

EXPERIMENTAL MORTALITY: The loss of experimental subjects occurring during an experiment because of moving away, falling sick or some other reason.

FORMATIVE EVALUATION: Evaluation undertaken during the formative stages of a project to provide planners with feedback so that they can alter and improve the programme.

FUNCTIONAL LITERACY: The process enabling a person to become not only literate, but also capable of performing adult tasks within his/her socio-economic and political setting.

HISTORY: Any event occurring between two chronological measurements in an experiment which may affect the results of the second measurement.

INDICATOR: Something that indicates, points, signifies; a gauge that represents another entity. Thus a high drop-out rate in an adult education programme may be an indicator of a lack of community motivation.

INPUT EVALUATION: Assessing the various resources used in conducting a programme.

INSTRUMENTATION: Preparing tools for the measurement of individual and social characteristics.

INTERVAL SCALE: A measurement scheme that rank orders the elements within a defined category assuming an equal distance between each element within the category.

LITERACY: The quality or state of being able to read and write day-to-day messages in the mother tongue.

LONGITUDINAL: Relating to length. Longitudinal evaluation studies measure the same subject at various times during a long period, sometimes of years.

MATURATION: Changes that may occur in a person or process over a period of time during the experiment, thereby influencing results.

MEAN: The sum of a group of scores divided by the number of scores.

MEASURE: To regulate by a standard; to estimate or appraise by a criterion which is usually predetermined. Also the score which is thus obtained.

MEASUREMENT: The assignment of numbers to objects or events according to rules, thereby making quantitatively precise statements.

MEASURES OF CENTRAL TENDENCY: Measures that identify the average of a group as well as the deviations of individual scores from that average. There are three such measures: median, mean and mode.

MEDIAN: The score in a group of scores that is midway in the distribution.

META-EVALUATION: Meta is linked with the name of a discipline to designate a related discipline designed to deal critically with the original one. Thus meta-evaluation is the evaluation of evaluation.

MILIEU: The setting or environment in which a person functions.

MODE: The score in a group of scores that occurs most often.

MODEL: A design, description or analogy used to help visualize or make understandable something that is more complex.

MODERNIZATION: The process by which economic, political, or social dimensions of a society may be changed to bring about a 'modern' society.

NOMINAL SCALE: A word, usually a number, used to identify a category that has nothing to do with the value of the category, e.g., 0 = male; 1 = female.

OBSERVABLE BEHAVIOUR: A mode of acting that can be viewed, coded and assessed, e.g., a particular manner of questioning used by a teacher.

ORDINAL SCALE: A classification scheme that categorizes and rank orders the elements within a classification group in relation to one another, e.g., the classification of teachers into I, II and III according to the number of students they teach.

OUTPUT EVALUATION: Assessing the quality and quantity of the final product(s) of the programme, also taking into account any unintended by-products of the programme.

PARADIGM: An example or pattern; a very clear example of an archetype. In evaluation, a paradigm is equivalent to the 'intellectual ideology' of an evaluator.

PARAMETER: Any one of a set of properties whose value determines the characteristics or behaviour of something.

POLITICIZATION: The process of becoming more aware of politics, political processes and/or actions, practices or policies which are a result of political decisions.

PRAXIS: The union of action and reflection by which individuals may transform the realities of the world in which they live.

PROCESS EVALUATION: Assessing procedural strategies and comparing effectiveness of different approaches to instruction, extension, animation and organization.

RANDOM SAMPLE: A representative portion chosen from among the population; each individual in the population has an equal chance of being selected each time a selection is made.

RATIO SCALE: A measurement scheme that rank orders within a defined category (nominal and ordinal scales) offers equal distances between the units of measurement (interval scale) and also has an *absolute zero*.

REGRESSION: A functional relationship between two or more elements that can be used to determine the value of other variables.

STANDARDIZED TESTS: Tests, the results of which can be compared with predetermined criteria and norms.

STATISTICAL ANALYSIS: An examination of complex relationships between variables using empirical data and rules of statistics.

STATISTICS: The science of methods for analysing data obtained from empirical observations to make descriptions or inferences.

STRATIFIED SAMPLE: A selected portion of the population where each stratum from the total population is appropriately represented.

SUB-CULTURE: A group, community or movement within a society that represents a variation of behaviour and ideals, often reflected in a change of dress, language or artefacts, as well as values.

SUMMATIVE EVALUATION: Assessment of the impact of the total product, programme, etc., comparing observed effects with anticipated or desired effects.

SYSTEM: A whole emerging from an interacting and interdependent set of parts, subject to a common plan and having a common purpose.

SYSTEMS MODEL: A model that looks at social reality as a system that can be described in terms of inputs, processes, outputs and context. (See also model and system.)

TAXONOMY: An orderly classification that has some theoretical underpinnings.

TRANSACTIONAL EVALUATION: The process of determining effects on and within a system as well as the end-results of the system.

UNESCO: United Nations Educational, Scientific, and Cultural Organization.

UNOBTRUSIVE OBSERVATIONS: Methods of examination in which the evaluators do not materially interfere in the situation, but rely on indirect procedures to gather data.

VOLUNTARISM: The principle or practice of reliance on voluntary action, not coercion.